

Digitized by the Internet Archive
in 2011 with funding from
Boston Library Consortium Member Libraries

<http://www.archive.org/details/inferencefordist00cher>

31
415
0.02-
20

DEWEY

Massachusetts Institute of Technology
Department of Economics
Working Paper Series

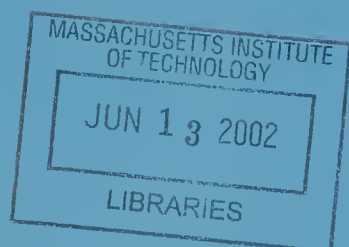
**INFERENCE FOR DISTRIBUTIONAL EFFECTS USING
INSTRUMENTAL QUANTILE REGRESSION**

Victor Chernozhukov, MIT
Christian Hansen, MIT

Working Paper 02-20
May 16, 2002

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

This paper can be downloaded without charge from the
Social Science Research Network Paper Collection at
http://papers.ssrn.com/paper.taf?abstract_id=313426



Massachusetts Institute of Technology
Department of Economics
Working Paper Series

**INFERENCE FOR DISTRIBUTIONAL EFFECTS USING
INSTRUMENTAL QUANTILE REGRESSION**

Victor Chernozhukov, MIT
Christian Hansen, MIT

Working Paper 02-20
May 16, 2002

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

This paper can be downloaded without charge from the
Social Science Research Network Paper Collection at
http://papers.ssrn.com/paper.taf?abstract_id=xxxxx

INFERENCE FOR DISTRIBUTIONAL EFFECTS USING INSTRUMENTAL QUANTILE REGRESSION

VICTOR CHERNOZHUKOV AND CHRISTIAN HANSEN

ABSTRACT. In this paper we describe how quantile regression can be used to evaluate the impact of treatment on the entire distribution of outcomes, when the treatment is endogenous or selected in relation to potential outcomes. We describe an instrumental variable quantile regression process and the set of inferences derived from it, focusing on tests of distributional equality, non-constant treatment effects, conditional dominance, and exogeneity. The inference, which is subject to the Durbin problem, is handled via a method of score resampling. The approach is illustrated with a classical supply-demand and a schooling example. Results from both models demonstrate substantial treatment heterogeneity and serve to illustrate the rich variety of hypotheses that can be tested using inference on the instrumental quantile regression process.

KEY WORDS: Quantile Regression, Instrumental Quantile Regression, Treatment Effects, Endogeneity, Stochastic Dominance, Hausman Test, Supply-Demand Equations, Returns to Education

Address correspondence to V. Chernozhukov, Department of Economics, MIT, E52-262F, 50 Memorial Drive, Cambridge, MA 02142 (e-mail vchern@mit.edu.) This is a provisional draft and an MIT Department of Economics Working Paper. March 2002. We thank the participants and organizers of the *EC*² 2001 meetings in Louvain-La-Neuve, where the research reported in this paper was presented.

1. INTRODUCTION

In this paper we describe how quantile regression can be used to evaluate the impact of treatment on the entire distribution of outcomes, when the treatment is self-selected or selected in relation to potential outcomes. We introduce an instrumental variable quantile regression process and the set of inferences derived from it, focusing on tests of distributional equality, non-constant treatment effects, conditional dominance, and exogeneity. The inference, which is subject to the Durbin problem, is handled via a method of score resampling. The paper describes these methods and establishes their asymptotic validity. The approach is illustrated with a classical supply-demand and a schooling example.

Inference about distributional outcomes is crucial in a wide range of economic analyses. For example, formal evaluation of a social program requires inference concerning the nature, direction, and magnitude of the program's impact throughout the entire outcome distribution, since evaluation involves integration of utility functions under alternative distributions of outcomes. For further examples and previous literature, see Atkinson (1970), Abadie (2002), Foster and Shorrocks (1988), Heckman and Smith (1997), and McFadden (1989). Just as in classical p -sample theory, e.g. Doksum (1974) and Shorack and Wellner (1986), this kind of inference is based on the empirical quantile regression process, which has recently been explored by Koenker and Xiao (2002).

The goal of this paper is to offer an empirical instrumental variable quantile regression (IV-QR) process and a set of inference tools derived from it. Effectively, instrumentation eliminates the endogeneity and selection bias commonly occurring in observational studies and experiments with imperfect compliance. Thus, the IV-QR process allows us to measure the exogenous treatment effect as in a fully controlled experiment, whereas the conventional QR process is inherently biased.

Using the instrumental variable quantile regression process we describe and derive the properties of a class of tests based on it which allow us to examine:

1. The hypothesis of *distributional equality*, or whether the treatment has a significant effect,
2. The hypothesis of non-constant or *varying treatment effect*, a fundamental hypothesis of causal analysis, cf. Heckman (1990), Doksum (1974), and Koenker and Xiao (2002),
3. The hypothesis of conditional *stochastic dominance*, a fundamental hypothesis as well, cf. Abadie (2002) and McFadden (1989),
4. The hypothesis of *exogeneity*, or whether the treatment variable is exogenous, another essential hypothesis, e.g. Hausman (1978).

A difficulty which arises when implementing these tests is that some of them are subject to the Durbin problem.¹ That is, the model's features or estimated nuisance parameters induce parameter dependent asymptotics, endangering distribution-free inference. A method of score resampling, which bootstraps the scores or estimated influence functions without recomputing the estimates, is suggested for generating asymptotically valid critical values for these tests. A main advantage of this method is its computational simplicity, which enables fast, practical implementation. The method is of independent interest in many other settings, and its immediate applicability to other problems is assured by the general conditions given in this paper. For example, it can be used for conventional quantile regression.

The use of the approach is illustrated through two empirical examples. In the first, we analyze the structure of demand for fish within a simultaneous equations demand system with random coefficients, and in the second, we consider the effect of schooling on earnings. We obtain clear evidence against the exogeneity and constant effect hypotheses, while accepting the hypothesis of first order stochastic dominance in both of these examples.

Thus, our paper complements the previous work which has considered research in this direction, including that of Hogg (1975), Abadie (1995) (2002), MaCurdy and Timmins (1998), McFadden (1989), Hong and Tamer (2001), Abadie, Angrist, and Imbens (2001), Koenker and Xiao (2002), Chernozhukov (2002) and others.

This paper accompanies our previous paper, Chernozhukov and Hansen (2001), that focused on modeling and identification of quantile treatment effects in the presence of endogeneity. The present paper goes further to establish the sampling properties of the entire instrumental variable quantile regression process and of the inference processes and test statistics derived from it. It also provides practical bootstrap tools to carry out the tests.

The remainder of the paper is organized as follows. In the next section, we briefly discuss the causal model and its examples that underlie both the estimation and the empirical examples developed in this paper. Section 3 then presents the IV-QR process and develops its sampling theory. Section 4 develops inference procedures for the IV-QR process and presents practical inference for testing distributional hypotheses. The use of the methods are illustrated via two empirical examples in Section 5, and Section 6 concludes.

2. THE CAUSAL MODEL

The following model is a simultaneous equations model. To describe it we will use a conventional potential outcomes framework.² *Potential or counterfactual* real-valued *outcomes* are indexed against treatment D ($D \in \mathcal{D}$, a subset of \mathbb{R}^J), and denoted Y_d , while potential

¹The term was coined by Koenker and Xiao (2002) to emphasize Durbin's contribution to theory of goodness-of-fit tests with estimated parameters and have an easy way to refer to the problem.

²See e.g. Heckman and Robb (1986) and Imbens and Angrist (1994)).

treatment status is indexed against the instrument Z , and denoted D_z . For example, Y_d is an individual's outcome when $D = d$ and D_z is an individual's treatment status when $Z = z$.

The potential or counterfactual outcomes $\{Y_d, d \in \mathcal{D}\}$, such as wages or demand, vary across individuals or states of the world. Given the actual treatment D , the observed outcome is

$$Y \equiv Y_D.$$

That is, only the D -th component of $\{Y_d, d \in \mathcal{D}\}$ is observed. Typically D is selected in relation to potential outcomes, *inducing endogeneity or sample selectivity*.

The objective of our analysis is to learn about the marginal distributions or, equivalently, the *conditional quantiles* of potential outcomes Y_d :

$$Q_{Y_d|X}(\tau), \tau \in \mathcal{T},$$

where \mathcal{T} is a closed subinterval of $(0, 1)$.

Quantiles of potential outcomes are a primary input to decisions about the efficiency of treatment and social programs³. The main obstacle to learning about the quantiles of potential outcomes is sample selectivity or endogeneity—the observed (Y_D, D) are typically misleading about the quantities in question, see e.g. Heckman (1990).

2.1. The Model. The following model has been suggested in Chernozhukov and Hansen (2001). This model rationalizes a Wald-type estimating equation and justifies a large variety of estimators based on it. In this model, the selection of treatment D by the individuals is left essentially unrestricted. It is assumed that there exists a vector of instrumental variables Z that affect the selection of D but do not affect the potential outcomes. Observed individual characteristics are denoted by X .

The main restriction of the model is the similarity assumption. The similarity assumption states that conditional on the information that led to an individual's selection of treatment state, the expectation of any function of the individual's rank does not vary across treatments. In other words, the selection presumes that the ex-ante expectation of a person's "ability", as measured by the person's rank in the distribution of unobservables, U_d in Assumption A1, relative to people with the same observed characteristics (X, Z) , does not vary across the treatments.

Assumption 1. (IV-QR Model) For almost every value of $(X, Z) = (x, z)$,

A1 POTENTIAL OUTCOMES. Given $X = x$, for some $U_d \stackrel{d}{\sim} U(0, 1)$,

$$Y_d \equiv q(d, x, U_d),$$

implying that $q(d, x, \tau)$ is the τ -th quantile of Y_d .

³This has been discussed, for example, in Abadie (2002).

A2 SELECTION. For unknown function $\delta(\cdot)$ and random vector V , the potential treatment indexed by instrument status z , given $X = x$, takes the form

$$D_z \equiv \delta(z, x, V).$$

A3 INDEPENDENCE. Given X ,

$$\{U_d\} \text{ is independent of } Z.$$

A4 SIMILARITY. For each d, d' , given (V, X, Z)

$$U_d \text{ is equal in distribution to } U_{d'}.$$

A5 OBSERVABLES consist of, for $U \equiv U_D$

$$\begin{cases} Y \equiv q(D, X, U), \\ D \equiv \delta(Z, X, V), \\ X, Z \end{cases}$$

It is important to note that the model differs from the conventional selection model of Heckman and Angrist and Imben's LATE. The differences mainly relate to the similarity assumption and slightly different independence conditions. An extensive comparative discussion of the IV-QR model is given in Chernozhukov and Hansen (2001). The role of each assumption is perhaps best highlighted in the demand and schooling examples described in the next two sections. The discussion is elaborate since these examples underlie the empirical analysis in this paper.

2.2. Example: A Demand Model. The following is a general simultaneous equation model, and many classical structural models can be written as special cases. Consider the following model

$$\begin{cases} \text{i.} & Y_p = q(p, U) & \text{demand,} \\ \text{ii.} & \tilde{Y}_p = \rho(p, z, \mathcal{U}) & \text{supply,} \\ \text{iii.} & P \in \{p : q(p, Z, U) = \rho(p, \mathcal{U})\} & \text{equilibrium.} \end{cases} \quad (2.1)$$

The map $p \mapsto Y_p$ is the random demand function, that is it is the demand when the price is p . Likewise, $p \mapsto \tilde{Y}_p$ is the random supply function, that is the supply when the price is p . Additionally, Y_p and \tilde{Y}_p , $q(\cdot)$, and $\rho(\cdot)$ depend on the covariates X , but this dependence is suppressed. Random variable U is the level of the demand in the sense that

$$q(p, U) \leq q(p, U') \text{ when } U \leq U'.$$

Demand is maximal when $U = 1$ and minimal when $U = 0$, holding p fixed. Likewise, \mathcal{U} is the level of supply. The τ -quantile of the demand curve $p \mapsto Y_p$ is given by

$$p \mapsto Q_{Y_p}(\tau) \equiv q(p, \tau).$$

Thus with probability τ , the curve $p \mapsto Y_p$ lies below the curve $p \mapsto Q_{Y_p}(\tau)$.

The quantile treatment effect is characterized by an elasticity

$$q(p', \tau) - q(p, \tau) \quad \text{or, if defined, by} \quad \partial \ln q(p, \tau) / \partial \ln p.$$

The elasticity depends on the state of the demand τ (low or high) and may vary considerably with τ . For example, this variation could arise when the number of buyers varies and aggregation induces non-constant elasticity across the demand levels as a process of summation of individual demand curves, holding the price fixed.

This example incorporates traditional models with additive errors

$$Y_p = q(p) + \varepsilon, \quad \text{where } \varepsilon = Q_\varepsilon(U). \quad (2.2)$$

Note that the model of demand in i. is more general in that the price elasticity is random, while in (2.2) it is constant. In other words, (2.2) restricts the price effect to parallel shifts in demand, while (2.1) allows for general, non-parallel effects.

Condition iii. is the equilibrium condition that generates *endogeneity* – the *selection* of the actual price by the market depends on the potential demand and supply outcomes i. and ii. As a result

$$P = \delta(Z, V),$$

where V consists of U , \mathcal{U} , and other variables (including “sunspot” variables, if the equilibrium price is not unique).

Instrumental variables Z , like *weather conditions* and *factor prices*, that shift the supply curve and do not affect the level of the demand curve U allow identification of the τ -th quantile of the demand function, $p \mapsto q(p, \tau)$. Perhaps remarkably, the model allows correlation between Z and V to exist.

2.3. Example: A Roy type Model. An individual considers two levels of schooling denoted $d = 0, 1$. The potential outcome under each schooling level is given by

$$\{Y_d, d = 0, 1\}.$$

The individual selects his schooling level to maximize his expected utility:

$$\begin{aligned} D &= \arg \max_{d \in \{0, 1\}} \left\{ E[W(Y_d) | X, Z, V] \right\} \\ &= \arg \max_{d \in \{0, 1\}} \left\{ E[W(q(d, X, U_d)) | X, Z, V] \right\} \end{aligned} \quad (2.3)$$

where W is the unobserved Bernoulli utility function, and E is the rational expectation.

As a result,

$$D = \delta(Z, X, V)$$

where Z, X are observed, V is an error vector that depends on ranks $\{U_d\}$ and other unobserved variables that affect the selection, and δ is an unknown function.

The *similarity assumption* imposes that

$$E[W(q(d, X, U_d)) | X, Z, V] = E[W(q(d, X, U_0)) | X, Z, V] \quad (2.4)$$

for any function W , where E is the rational expectation (computed with respect to the true probability law P). In other words, the decision maker's information does not allow the objective discrimination of systematic variation of his ranks across the treatment states, where the *ranks are defined relative to the observationally identical group of individuals with the same X and Z* . In other words, the similarity assumption is nothing but a restriction on the information set of the individual.

Note that similarity is defined relative to the true expectation. In effect, (2.4) does not require an individual to have rational expectations while optimizing in (2.3). Indeed, we could replace the expectation operator E in formula (2.3) by some subjective expectations E_B such that

$$D = \arg \max_{d \in \{0,1\}} \left\{ E_B[W(q(d, X, U_d)) | X, Z, V] \right\}.$$

3. THE INSTRUMENTAL VARIABLE QUANTILE REGRESSION PROCESS

3.1. The Principle. We first describe two important implications that arise from assumptions A1-A4.

Proposition 1 (A1-A5 imply Wald IV). *Suppose A1-A5 and that $\tau \mapsto q(D, X, \tau)$ is continuous and strictly increasing a.s. Then for each $\tau \in (0, 1)$*

$$P[Y < q(D, X, \tau) | X, Z] = \tau, \text{ a.s.} \quad (3.1)$$

Furthermore,

$$0 = Q_{Y - q(D, X, \tau) | X, Z}(\tau) \quad \text{a.s.} \quad (3.2)$$

Equation (3.1) is a Wald style restriction that can be used to estimate the quantile process $\tau \mapsto q(d, x, \tau)$. Identification of the quantile process in the population does not require functional form assumptions. This result is reported in Chernozhukov and Hansen (2001).

The main identification restriction (3.1) can be posed via (3.2) as an optimization problem, which we call the *instrumental variable* or *inverse quantile regression* for its “inverse” relation to the (conventional) quantile regression of Koenker and Bassett (1978). This links the IV-QR model and quantile regression together.

Recall from Koenker and Bassett (1978) that quantile regression is formulated as finding the best predictor of Y given X under the expected loss using the asymmetric least absolute deviation criterion:

$$\rho_\tau(u) = \tau u^+ + (1 - \tau) u^-.$$

In other words, assuming integrability, the τ -th conditional quantile of Y given X solves:

$$Q_{Y|X}(\tau) = \arg \min_f E[\rho_\tau(Y - f(X))|X].$$

Note that median regression is one important instance of quantile regression.

Proposition 1 states that 0 is the τ -th quantile of random variable $Y - q(D, X, \tau)$ conditional on (X, Z) :

$$0 = Q_{Y-q(D,X,\tau)}(\tau|X, Z) \quad \text{a.s. for each } \tau.$$

Thus, we may pose the problem of finding a function $q(d, x, \tau)$ satisfying equation (3.2) of Proposition 1 as the *instrumental variable* or *inverse quantile regression*:

Find a function $q(x, d, \tau)$ such that $\mathbf{0}$ is a solution to the quantile regression of $Y - q(X, D, \tau)$ on (Z, X) :

$$0 = \arg \min_f E\rho_\tau[(Y - q(D, X, \tau) - f(X, Z))|X, Z].$$

3.2. An Instrumental Variable Quantile Regression Process. For estimation of the IV-QR process, we focus on the basic linear model, which covers a wide range of applications:

$$q(D, X, \tau) = D'\alpha(\tau) + X'\beta(\tau), \quad (3.3)$$

where D is an l -vector of treatment variables (possibly interacted with covariates) and X is a k -vector of (transformations) of covariates. The linear quantile model is obviously a special case of the more general model presented in Assumption A1, and is a basic model of quantile regression research.

We focus on a simple finite-sample analog of the the instrumental variable quantile regression in the population. Define the weighted quantile regression objective function as

$$Q_n(\tau, \alpha, \beta, \gamma) \equiv \frac{1}{n} \sum_{i=1}^n \rho_\tau(Y_i - D_i'\alpha - X_i'\beta - \hat{\Phi}_i(\tau)'\gamma) \cdot \hat{V}_i(\tau), \quad \text{where}$$

$\Phi_i(\tau) \equiv \Phi(\tau, X_i, Z_i)$ is an l -vector of instruments ($l = \dim(D)$),

$V_i(\tau) \equiv V(\tau, X_i, Z_i) > 0$ is a weight function.

In principle, we may consider a larger number of elements in vector Φ . However, this is not necessary as efficiency can instead be achieved by choosing Φ and V appropriately.

The IV-QR procedure is as follows: for $\|x\|_A = \sqrt{x'Ax}$

$$\hat{\alpha}(\tau) = \arg \inf_{\alpha \in \mathcal{A}} \|\hat{\gamma}(\alpha, \tau)\|_A, \quad \text{such that} \quad (3.4)$$

$$(\hat{\beta}(\alpha, \tau), \hat{\gamma}(\alpha, \tau)) = \arg \inf_{(\beta, \gamma) \in \mathcal{B} \times \mathcal{G}} Q_n(\tau, \alpha, \beta, \gamma), \quad (3.5)$$

where \mathcal{A} and \mathcal{B} are parameter sets, \mathcal{G} is any fixed compact cube centered at 0, and A is a positive definite matrix.

The parameter estimates are given by:

$$\left(\hat{\alpha}(\tau), \hat{\beta}(\tau)\right) \equiv \left(\hat{\alpha}(\tau), \hat{\beta}(\hat{\alpha}(\tau), \tau)\right) \quad (3.6)$$

Let us denote

$$\hat{\theta}(\tau) = (\hat{\alpha}(\tau)', \hat{\beta}(\tau)')' \text{ and } \theta(\tau) = (\alpha(\tau)', \beta(\tau)')'.$$

There are three principal motivations for this estimator. First, it naturally links IV restrictions and conventional quantile regression together by exploiting the principle described in the previous section. Second, it is computationally convenient, since the estimates can be computed by implementing a series of ordinary quantile regressions (convex optimization problems) implying a need for a grid search only over the α -parameter. Third, it is asymptotically equivalent to a GMM estimator and achieves maximal efficiency by choosing instruments Φ and weights V appropriately.

The *instrumental variable quantile regression process* is defined as

$$\hat{\theta}(\cdot) \equiv \left(\hat{\theta}(\tau), \tau \in \mathcal{T}\right),$$

where \mathcal{T} is a closed subinterval of $(0, 1)$.

3.3. Assumptions. In order to obtain properties of the IV-QR process, we first impose a set of simple regularity conditions.

Assumption 2 (Conditions for Estimation). In addition to (3.3), suppose

- R1 SAMPLING.** (Y_i, D_i, X_i, Z_i) are iid defined on the probability space (Ω, \mathcal{F}, P) and take values in a compact set.
- R2 COMPACTNESS.** For all τ , $(\alpha(\tau), \beta(\tau)) \in \text{int } \mathcal{A} \times \mathcal{B}$, $\mathcal{A} \times \mathcal{B}$ is compact and convex.
- R3 FULL RANK AND CONTINUITY.** a.s. $\sup_{y \in \mathbb{R}} f_{Y|(X,D,Z)}(y) < K$ and

$$\frac{\partial}{\partial(\alpha', \beta', \gamma')} E_P [1(Y < D'\alpha + X'\beta + \Phi(\tau)'\gamma)\Psi(\tau)], \quad \Psi_i(\tau) \equiv V_i(\tau) \cdot [X_i' : \Phi_i(\tau)']'$$

has full rank, and is continuous in $(\alpha, \beta, \gamma, \tau)$ uniformly over $\mathcal{A} \times \mathcal{B} \times \mathcal{G} \times \mathcal{T}$.

- R4 ESTIMATED INSTRUMENTS AND WEIGHTS.** Wp $\rightarrow 1$, $\hat{\Phi}(\tau, z, x), \hat{V}(\tau, z, x) \in \mathcal{F}$ and $\hat{V}(\tau, z, x) \xrightarrow{p} V(\tau, z, x) \in \mathcal{F}$, $\hat{\Phi}(\tau, z, x) \xrightarrow{p} \Phi(\tau, z, x) \in \mathcal{F}$ uniformly in (τ, z, x) over compact sets. For all τ , functions $\mathcal{F}: (\tau, z, x) \mapsto f(\tau, z, x)$ are uniformly smooth functions in (z, x) , C_K^η ,⁴ with the uniform smoothness order $\eta > \dim(d, z, x)/2$, and f are uniformly Holder in τ : $\|f(\tau', z, x) - f(\tau, z, x)\| < C|\tau - \tau'|^a$, where the constants $C > 0$, and $0 < a \leq 1$ are independent of (z, x, τ, τ') .

⁴See page 154 in van der Vaart and Wellner (1996).

Condition R1 imposes iid sampling and compactness on the support of the economic variables. The compactness is hardly restrictive in micro-econometric applications, but it can be relaxed. Condition R2 imposes compactness on the parameter space. Such an assumption is needed at least for the parameter $\alpha(\tau)$ since the objective function is not convex in α . The full rank condition in R3 implies global identification, and the continuity condition in R3 together with R1 suffices for asymptotic normality. The parametric identification condition is similar in spirit to the nonparametric identification conditions obtained by Chernozhukov and Hansen (2001) and is of independent interest. Essentially, this condition requires that the instrument Φ impacts the joint distribution of (Y, D) at many relevant points. Clearly, conditions R1-R4 may be refined at a cost of more complicated notation and proof.

The role of R4 is to allow possibly estimated instruments and weights. Smoothness in R4 needs to hold only for the non-discrete sub-component of (d, x, z) . Condition R4 allows for a wide variety of nonparametric and parametric estimators, as shown by Andrews (1994). See also Andrews (1995), Newey (1990, 1997), and Newey and Powell (1990) for other examples of such estimators. The smoothness condition in R4 can be replaced by a more general condition of \mathcal{F} having a finite $L_2(P)$ -bracketing entropy integral.

3.4. Estimation Theory. Theorem 1 describes the distribution of the IV-QR process.

Theorem 1 (IV-QR Process). *Given Assumptions 1-2, for $\epsilon_i(\tau) = Y_i - D_i'\alpha(\tau) + X_i'\beta(\tau)$*

$$\begin{aligned}\sqrt{n}(\hat{\theta}(\tau) - \theta(\tau)) &= J(\tau)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n l_i(\tau, \theta(\tau)) \Psi_i(\tau) + o_p(1) \\ &\Rightarrow b(\tau) \text{ in } \ell^\infty(\mathcal{T}),\end{aligned}\tag{3.7}$$

uniformly in τ , where

$$l_i(\tau, \theta(\tau)) = (\tau - 1(\epsilon_i(\tau) < 0))$$

and $b(\cdot)$ is a mean zero Gaussian process with covariance function:

$$E b(\tau)b(\tau') = J(\tau)^{-1} S(\tau, \tau') [J(\tau')^{-1}]',$$

where

$$J(\tau) = E [f_{\epsilon(\tau)}(0|X, D, Z) \Psi(\tau) [D' : X']], \quad S(\tau, \tau') = (\min(\tau, \tau') - \tau\tau') E \Psi(\tau) \Psi(\tau')'.$$

Theorem 1 simply states that $\hat{\theta}(\cdot)$ is approximately distributed as a continuous random Gaussian function. This implies a variety of useful results.

Corollary 1 (Normality). *For any finite collection of quantile indices $\{\tau_j, j \in J\}$*

$$\left\{ \sqrt{n} \left(\hat{\theta}(\tau_j) - \theta(\tau_j) \right) \right\}_{j \in J} \xrightarrow{d} N \left(0, \left\{ J(\tau_k)^{-1} S(\tau_k, \tau_l) [J(\tau_l)^{-1}]' \right\}_{k, l \in J} \right).$$

Corollary 2 (Efficient Weights and Instruments). *When we choose the weights $V^*(\tau) = f_{\epsilon(\tau)}(0|X, Z)$, $v(\tau) = f_{\epsilon(\tau)}(0|D, X, Z)$, $\Phi^*(\tau) = E[Dv(\tau)|X, Z]/V^*(\tau)$, and $\Psi^*(\tau) = V^*(\tau)[X' : \Phi^*(\tau)]$, the covariance function equals*

$$E b(\tau)b(\tau')' = (\min(\tau, \tau') - \tau\tau') \cdot [\Psi^*(\tau)\Psi^*(\tau')']^{-1}.$$

This choice of instruments and weights leads to an efficient procedure. This can be shown by appealing to Chamberlain's (1986) arguments. Regularity condition R4 allows use of a wide variety of nonparametric estimators and parametric approximations of the optimal Φ and V . For particular examples of such procedures, see Amemiya (1977), Andrews (1994, 1995), Newey (1990, 1997), and Newey and Powell (1990). An example of a simple and practical strategy for empirical work is to construct Φ as an OLS projection of D on X and Z (and possibly their powers) and set $V_i = 1$.

Corollary 3 (Distribution-Free Limits). *For $W(\tau) = \tau(1-\tau)J(\tau)^{-1}E\Psi(\tau)\Psi(\tau)'[J(\tau)^{-1}]'$*

$$W(\tau)^{-\frac{1}{2}}\sqrt{n}\left(\hat{\theta}(\tau) - \theta(\tau)\right) \Rightarrow B_p(\tau),$$

where B_p is a standard p -dimensional Brownian bridge B_p ($p = l + k$) with covariance operator

$$E B_p(\tau)B_p(\tau') \equiv (\min(\tau, \tau') - \tau\tau')I_p.$$

4. INFERENCE

Several distributional hypotheses have been posed in the fundamental econometric and statistical literature. For example, an essential hypothesis is whether the treatment exhibits a pure location (constant treatment) effect or a general shape effect, e.g. Doksum (1974) and Koenker and Xiao (2002), or whether the treatment creates a stochastic dominance effects, cf. Abadie (2002), Heckman and Smith (1997), and McFadden (1989). In structural and causal empirical analysis, the hypothesis of endogeneity is also fundamental, motivating Hausman tests, cf. Hausman (1978). In this section, we describe inference procedures to test these hypotheses.

4.1. The Inference Problem. All of our hypotheses will be embedded in the following null hypothesis:

$$R(\tau)\left(\theta_n(\tau) - r_n(\tau)\right) = 0, \quad \text{for each } \tau \in \mathcal{T}, \quad (4.1)$$

where $R(\tau)$ denotes a known $q \times p$ matrix, $q \leq p = \dim(\theta)$ and $r \in \mathbb{R}^p$. The parameter $\theta_n(\cdot)$ is made explicitly dependent on the sample size n to accomodate asymptotic analysis under local alternatives.

The tests will be based on the instrumental variable quantile regression process, $\hat{\theta}(\cdot)$. We will focus on the basic inference process,

$$v_n(\tau) = R(\tau) \left(\hat{\theta}_n(\tau) - \hat{r}_n(\tau) \right), \quad (4.2)$$

and statistics of the form $S_n = f(\sqrt{n}v_n(\cdot))$ derived from it. In particular, we will be interested in the Kolmogorov-Smirnov (KS) and Smirnov-Cramer-Von-Misses (CM) statistics, which have

$$S_n = \sqrt{n} \sup_{\tau \in \mathcal{T}} \|v_n(\tau)\|_{\hat{\Lambda}(\tau)}, \quad S_n = n \int_{\mathcal{T}} \|v_n(\tau)\|_{\hat{\Lambda}(\tau)}^2 d\tau, \quad (4.3)$$

respectively, where $\|a\|_{\Lambda} \equiv \sqrt{a' \Lambda a}$, the symmetric $\hat{\Lambda}(\tau) \xrightarrow{p} \Lambda(\tau)$ uniformly in τ , and $\Lambda(\tau)$ is a positive definite symmetric matrix uniformly in τ . The choice of Λ and $\hat{\Lambda}$ is discussed in Sections 4.3 and 4.5.

Example 1 (*Hypothesis of Equality of Distributions*). A basic hypothesis is that the treatment impacts the outcomes significantly:

$$\alpha(\tau) \neq 0 \text{ for some } \tau \text{ in } \mathcal{T}.$$

In this case,

$$R(\tau) = R = [1, 0, \dots] \text{ and } r(\tau) = 0.$$

The next example imposes a restrictive, yet simple mechanism through which the treatment may operate. This mechanism requires that the effect is constant across the distribution. The alternative is that the effect varies across quantiles. The alternative hypothesis of heterogeneous treatment effects is of fundamental importance because it motivates the modern causal models, see e.g. Heckman (1990), which were developed specifically to cope with varying effects.

Example 2 (*Location-Shift or Constant Effect Hypothesis*). The hypothesis of a *constant treatment effect* is that the treatment D affects only the location of outcome Y , but not any other moments. That is,

$$\exists \alpha : \alpha(\tau) = \alpha, \text{ for each } \tau \in \mathcal{T}.$$

In this case,

$$R(\tau) = R = [1, 0, \dots] \text{ and } r(\tau) = r = (\alpha, \beta)', \text{ implying} \\ Rr = \alpha,$$

which asserts that the $\alpha(\tau)$ is constant across all $\tau \in \mathcal{T}$. The component r can be estimated by any method consistent with the null, e.g. $\hat{r} = (\hat{\alpha}(\frac{1}{2})', \hat{\beta}(\frac{1}{2})')'$.

Example 3 (*Dominance Hypothesis*). The test of *stochastic dominance*, or whether the treatment is unambiguously beneficial, involves the dominance null

$$\alpha(\tau) \geq 0, \text{ for all } \tau \in \mathcal{T}$$

versus the non-dominance alternative

$$\alpha(\tau) < 0, \text{ for some } \tau \in \mathcal{T}.$$

In this case, the *least favorable null* involves

$$R(\tau) = R = [1, 0...] \text{ and } r(\tau) = 0,$$

and one may use the *one-sided* KS or CM statistics, cf. Abadie (2002),

$$S_n = \sqrt{n} \sup_{\tau \in \mathcal{T}} \max(-\hat{\alpha}(\tau), 0), \text{ and } S_n^- = \bar{n} \int_{\mathcal{T}} \|\max(-\hat{\alpha}(\tau), 0)\|_{V(\tau)}^2 d\tau$$

to test the hypothesis.

Example 4 (*Exogeneity Hypothesis*). In a basic model, the quantiles of potential or counterfactual outcomes, conditional on X , are given by

$$Q_{Y_d|X}(\tau) = d'\alpha(\tau) + X'\beta(\tau).$$

Suppose that the treatment D is chosen *independently of outcomes*, that is D is independent of $\{U_d\}$, conditional on X . Then the quantiles of realized outcome Y , conditional on D and X , are given by

$$Q_{Y|D,X}(\tau) = D'\alpha(\tau) + X'\beta(\tau).$$

Thus, in the absence of endogeneity, $(\alpha(\tau)', \beta(\tau)')'$ can be estimated using the conventional quantile regression without instrumenting. The difference between IV-QR estimates, $\hat{\theta}(\cdot)$, and QR estimates, $\hat{\vartheta}(\cdot)$, can be used to formulate a Hausman test of the null hypothesis of *no endogeneity*:

$$\alpha(\tau) = \vartheta(\tau)_1 \text{ for each } \tau \text{ in } \mathcal{T}, \text{ where } \vartheta(\tau) \equiv \text{plim } \hat{\vartheta}(\tau), \quad (4.4)$$

and $\hat{\vartheta}(\cdot)_1$ is the QR estimate of $\alpha(\cdot)$ obtained without instrumenting. In this case,

$$R(\tau) = [1, 0, ...], \text{ and } r(\tau) = \vartheta(\tau).$$

The alternative of endogeneity states:

$$\exists \tau \in \mathcal{T} : \alpha(\tau) \neq \vartheta(\tau)_1. \quad (4.5)$$

4.2. The Assumptions. We will maintain the following technical assumptions.

Assumption 3 (Conditions for Inference).

I.1 (Y_i, D_i, X_i, Z_i) are iid on the probability space (Ω, F, P_n) . The law of $(Y_t, D_t, Z_t, X_t, t \leq n)$, $P_n^{[n]}$, is contiguous to some $P^{[n]}$,⁵ and either

(a) for a fixed continuous function $p(\tau) : \mathcal{T} \rightarrow \mathbb{R}^q$ and for each n

$$R(\tau) (\theta_n(\tau) - r_n(\tau)) = g(\tau), \quad g(\tau) = p(\tau)/\sqrt{n}, \text{ or,}$$

(b) for a fixed continuous function $g(\tau) : \mathcal{T} \rightarrow \mathbb{R}^q$ and for each n

$$R(\tau) (\theta(\tau) - r(\tau)) = g(\tau).$$

⁵Contiguity is defined as on p. 87 in van der Vaart (1998)

Functions $R(\tau)$, $g(\tau)$, $p(\tau)$, and $\lim_n r_n(\tau)$ are continuous in τ .

I.2 (a) Under any local alternative, I2(a)

$$\sqrt{n}(\hat{\theta}(\cdot) - \theta_n(\cdot)) \Rightarrow b(\cdot),$$

$$\sqrt{n}(\hat{r}(\cdot) - r_n(\cdot)) \Rightarrow d(\cdot),$$

jointly in $\ell^\infty(\mathcal{T})$, where $b(\cdot)$ and $d(\cdot)$ are jointly zero mean Gaussian functions.

(b) Under the global alternative, I2(b), the same holds, except that the limit $(\tilde{b}(\cdot), \tilde{d}(\cdot))$ needs not have the same distribution as in I2(a).

Conditions I.1(a) and I.1(b) formulate a local and a global alternative.

Condition I.2 requires that the estimates of $\theta(\cdot)$ and $r(\cdot)$ are asymptotically Gaussian. In our examples, this is guaranteed by Theorem 1 and asymptotic results for conventional QR. Note that I.2 is formulated so that other asymptotically Gaussian estimators of the parameters of the IV-QR model are permitted. Detailed discussion of this assumption is stated after Assumption 4.

4.3. Inference Theory. We are now prepared to state a main result.

Theorem 2 (Inference). *For f denoting the two- and one- sided KS or CM statistics*

1. *Under Assumptions 1, 2, 3:I1(a), and 3:I2(a) in $\ell^\infty(\mathcal{T})$*

$$S_n \Rightarrow S \equiv f(v_0(\cdot) + p(\cdot)),$$

where $v_0(\tau) = R(\tau)(b(\tau) - d(\tau))$. If $v_0(\cdot)$ has non-generate covariance kernel, when the null is true ($p \equiv 0$), we have for $\alpha < 1/2$

$$P_n(S_n > c(1 - \alpha)) \rightarrow \alpha = P(f(v_0(\cdot)) > c(1 - \alpha)),$$

and when the null is not true ($p \neq 0$),

$$P_n(S_n > c(1 - \alpha)) \rightarrow \beta = P(f(v_0(\cdot) + p(\cdot)) > c(1 - \alpha)) > \alpha.$$

2. *Under Assumptions 1, 2, 3:I1(b), and 3:I2(b),*

$$S_n \xrightarrow{p_n} \infty, \quad P_n(S_n > c(1 - \alpha)) \rightarrow 1.$$

Theorem 2 states the limit distribution of the KS and CM statistics under local alternatives and the global alternative. Note that in the statement of Theorem 2 we implicitly assume that for the case of one-sided tests in Example 3, the local or global alternatives to the least favorable null violate the composite null.

As it stands, Theorem 2 does not provide us with operational tests, since we do not know the critical value

$$c(1 - \alpha) : P(f(v_0(\cdot)) > c(1 - \alpha)) = \alpha.$$

In general, one faces the Durbin problem when estimating $c(1 - \alpha)$ since the limit is generally not-distribution free and simulations are infeasible.

In several important cases, such as Examples 1 and 3, the Durbin component $d(\cdot)$ is equal to zero and so is not present. In these cases, it is possible by picking an appropriate weight matrix $\Lambda(\tau)$ to achieve asymptotically distribution-free inference, at least under iid sampling.

Corollary 4 (Distribution-Free Inference). *Suppose $d(\cdot) = 0$. If we pick matrix $\Lambda(\tau) = [R(\tau)W(\tau)R(\tau)]^{-1}$ with $W(\tau) = \tau(1 - \tau)J(\tau)^{-1}E[\Psi(\tau)\Psi(\tau)']J(\tau)^{-1}$ to enter the norm $\|\cdot\|_{\Lambda(\tau)}$ in the definition of the KS and CM statistics f in (4.3), and suppose we have $\hat{\Lambda}(\tau) = \Lambda(\tau) + o_p(1)$ uniformly in τ , then*

$$f(v_n(\cdot)) \Rightarrow f(B_q(\cdot)),$$

where B_q is the standard q -dimensional Brownian bridge with covariance function:

$$EB_q(\tau)B_q(\tau)' = (\min(\tau, \tau') - \tau\tau')I_q.$$

In other important cases, such as Examples 2 and 4, the transformation used in Corollary 4 will not provide distribution-free limits. There are several ways to proceed. One method is a martingale transformation using Khmaladization, cf. Koenker and Xiao (2002). Another method is a simple resampling with recentring the inference process around its sample realization, cf. Chernozhukov (2002). Simulation results in Chernozhukov (2002) suggest that resampling has an accurate size and somewhat better power than Khmaladization. In the next section, we describe a different resampling method that delivers the same asymptotic quality and is attractive computationally.

4.4. Inference by Resampling. The method of resampling we suggest in this paper does not require the recomputation of the estimates over the resampling steps, which may be quite laborious since the optimization problem requires many computations of ordinary quantile regressions for many values of α and τ . Instead we resample the linear approximation of the empirical inference processes. In addition, to facilitate a feasible, practical implementation, we employ the m out of n bootstrap (subsampling).⁶

Suppose that we have a linear representation for the inference process:

$$\sqrt{nv_n}(\cdot) - \sqrt{ng}(\cdot) = \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i(\cdot) + o_{p_n}(1), \quad (4.6)$$

where $z_i(\cdot)$ is defined below in Proposition 2.

Given a sample of the estimated scores (estimation is discussed below),

$$\{\hat{z}_i(\tau), i \leq n, \tau \in \mathcal{T}\},$$

consider the following steps.

⁶The full bootstrap is hardly feasible in our second empirical example, even though we are bootstrapping a linear statistic.

Step 1. Construct all subsets I of $\{j \leq n\}$ of size b . Denote such subsets as $I_i, i \leq B_n$. The number of such subsets B_n is “ n choose b .”⁷

Denote by $v_{j,b,n}$ the inference process computed over the j -th subset of data I_j ,

$$v_{j,b,n}(\tau) \equiv \frac{1}{b} \sum_{i \in I_j} \hat{z}_i(\tau),$$

and define $S_{j,b,n} \equiv f(\sqrt{b}[v_{j,b,n}(\cdot)])$ as

$$\hat{S}_{j,b,n} \equiv \sup_{\tau \in \mathcal{T}} \sqrt{b} \|v_{j,b,n}(\tau)\|_{\dot{V}(\tau)} \text{ or } \tilde{S}_{j,b,n} \equiv b \int_{\mathcal{T}} \|v_{j,b,n}(\tau)\|_{\dot{V}(\tau)}^2 d\tau,$$

for cases when S_n is the Kolomogorov-Smirnov (KS) or Smirnov-Cramer-Von-Misses (CM) statistic, respectively. Define for $S = f(v_0(\cdot))$

$$\Gamma(x) \equiv \Pr\{S \leq x\}.$$

Step 2. Estimate $\Gamma(x)$ by

$$\hat{\Gamma}_{b,n}(x) = B_n^{-1} \sum_{j=1}^{B_n} 1\{S_{j,b,n}(\tau) \leq x\}.$$

Step 3. The critical value is obtained as the $1 - \alpha$ -th quantile of $\hat{\Gamma}_{b,n}(\cdot)$:

$$c_{b,n}(1 - \alpha) = \hat{\Gamma}_{b,n}^{-1}(1 - \alpha) = \inf\{c : \Gamma_{b,n}(c) \geq 1 - \alpha\}.$$

The size α test rejects the null hypothesis when $S_n > c_{b,n}(1 - \alpha)$.

In obtaining the linear representation (4.6), we make use of the following assumption.

Assumption 4 (Linear Representations). In addition to I.1 and I.2

I.3 (a) Under any local alternative, I2(a), there exist sums of iid mean zero vectors such that uniformly in τ in \mathcal{T}

$$\sqrt{n} \left(\hat{\theta}(\cdot) - \theta_n(\cdot) \right) = J(\cdot)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n l_i(\cdot, \theta_n(\cdot)) \Psi_i(\cdot) + o_{p_n}(1) \Rightarrow b(\cdot),$$

$$\sqrt{n} (\hat{r}(\cdot) - r_n(\cdot)) = H(\cdot)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n d_i(\cdot, r_n(\cdot)) \Upsilon_i(\cdot) + o_{p_n}(1) \Rightarrow d(\cdot),$$

jointly in $\ell^\infty(\mathcal{T})$, where $b(\cdot)$ and $d(\cdot)$ are jointly zero mean Gaussian functions.

(b) Under the global alternative, I2(b), the same holds, except that the limit $(\tilde{b}(\cdot), \tilde{d}(\cdot))$ needs not have the same distribution as in I3(a).

⁷A smaller number B_n of randomly chosen subsets can also be used, if $B_n \rightarrow \infty$ as $n \rightarrow \infty$, cf. Section 2.5 in Politis, Romano, and Wolf (1999). The subsampling is done with replacement. However, if $b^2/n \rightarrow 0$, the subsampling without and subsampling with replacement are equivalent wp $\rightarrow 1$.

- I.4** (a) We have the estimates $\tau \mapsto l_i(\tau, \hat{\theta}(\tau))\hat{\Psi}_i(\tau)$ and $\tau \mapsto d_i(\tau, \hat{r}(\tau))\hat{\Upsilon}_i(\tau)$ that take realizations in a Donsker class of functions with a constant envelope and are uniformly consistent in τ under the $L_2(P)$ semimetric.⁸
- (b) $\text{Wp} \rightarrow 1$ for each i : $E_{P_n} l_i(\tau, \theta_n(\tau)) f_i(\tau) \big|_{f=\hat{\Psi}} = 0$, $E_{P_n} d_i(\tau, r_n(\tau)) f_i(\tau) \big|_{f=\hat{W}} = 0$.
- (c) Functions l_i and d_i are $L_2(P)$ -Lipschitz uniformly in τ : $[E_P \|l_i(\tau, \theta) - l_i(\tau, \theta')\|^2]^{1/2} < C \|\theta' - \theta\|$, $[E_P \|d_i(\tau, r) - d_i(\tau, r')\|^2]^{1/2} < C \|r' - r\|$, uniformly in (θ, θ', r, r') over compact sets.

Proposition 2 (Linear Representations). *Under Assumption 3-4 we have:*

$$\sqrt{n}v_n(\cdot) - \sqrt{n}g(\cdot) = \frac{1}{\sqrt{n}} \sum_{i=1}^n z_i(\cdot) + o_p(1), \quad (4.7)$$

where

$$z_i(\cdot) = R(\cdot) [J(\cdot)^{-1} l_i(\cdot, \theta(\cdot)) \Psi_i(\cdot) - H(\cdot)^{-1} d_i(\cdot, r(\cdot)) \Upsilon_i(\cdot)].$$

Thus, the estimate of $z_i(\cdot)$ is given by

$$\hat{z}_i(\tau) = R(\tau) [\hat{J}(\tau)^{-1} l_i(\tau, \hat{\theta}(\tau)) \hat{\Psi}_i(\tau) - \hat{H}(\tau)^{-1} d_i(\tau, \hat{r}(\tau)) \hat{\Upsilon}_i(\tau)],$$

where $\hat{J}(\tau)$ and $\hat{H}(\tau)$ are any uniformly consistent estimates.

In Assumption 4 condition I.3 requires that the estimates of θ and r entering the null hypotheses have asymptotically linear representation in the form defined above, and asymptotic normality applies to these estimates. Note that I.3 is formulated so that other asymptotically Gaussian estimators of the IV-QR model are permitted. Conditions I.4(a) and I.4(c) impose a smoothness needed for developing the theory of the resampling inference. These condition are also satisfied in all of the examples considered in this paper. Condition I.4(b) is the condition of “orthogonality” which implies that the estimation of weights Ψ_i and Υ_i has no effect on the asymptotic distribution of the linear representation in I.3.

Proposition 3 in Appendix C formally verifies that I.3 and I.4 are satisfied for the particular implementations that we propose. Next, we briefly go through examples and state the scores z_i for each of them.

1. Test of Equality of Distributions: Since $r = 0$ is not estimated,

$$z_i(\tau) = R(\tau) [J(\tau)^{-1} l_i(\tau, \theta(\tau)) \Psi_i(\tau)],$$

where $l_i(\tau, \theta(\tau)) = (\tau - 1(Y_i < D_i \alpha(\tau) + X_i' \beta(\tau)))$, $\Psi_i(\tau) = V_i(\tau) [\Phi_i(\tau)', X_i']'$

2. Test of Constant Effect: In this case, $\hat{r}(\cdot) = \hat{\theta}(\frac{1}{2})$ is an IV-QR estimate, and for $l_i(\cdot)$ defined above

$$z_i(\tau) = R(\tau) [J(\tau)^{-1} l_i(\tau, \theta(\tau)) \Psi_i(\tau) - J(\frac{1}{2})^{-1} l_i(\frac{1}{2}, \theta(\frac{1}{2})) \Psi_i(\frac{1}{2})],$$

⁸ $\hat{f}(W, \tau)$ is consistent to $f(W, \tau)$ under $L_2(P)$ if $\sup_{\tau} \text{Var} [\hat{f}(w, \tau) - f(w, \tau)]_{\hat{f}=f} \xrightarrow{P} 0$.

3. Test of Dominance Effect: Since $r = 0$, the score is

$$z_i(\tau) = R(\tau) [J(\tau)^{-1} l_i(\tau, \theta(\tau)) \Psi_i(\tau)],$$

4. Test of Exogeneity: If r is estimated using conventional quantile regression as defined in Example 4, the score is given by

$$z_i(\tau) = R(\tau) [J(\tau)^{-1} l_i(\tau, \theta(\tau)) \Psi_i(\tau) - H(\tau)^{-1} d_i(\tau, \vartheta(\tau))],$$

where $d_i(\tau, \vartheta(\tau)) = (\tau - 1(Y_i < \tilde{X}_i' \vartheta(\tau))) \tilde{X}_i$, $\tilde{X}_i = (D_i', X_i')'$ $H(\tau) = E f_{Y|\tilde{X}}(\vartheta(\tau)' \tilde{X}) \tilde{X} \tilde{X}'$.

Estimation of H and J matrices is further discussed in section 4.5.2. The next Theorem 3 establishes the properties of the proposed resampling method.

Theorem 3 (Resampling Inference). *Suppose Assumptions 2-3, and that we have $\hat{J}(\tau) \xrightarrow{p_n} J(\tau)$ and $\hat{H}(\tau) \xrightarrow{p_n} H(\tau)$ uniformly in τ over \mathcal{T} . Then as $b/n \rightarrow 0, b \rightarrow \infty, n \rightarrow \infty$*

1. *When the null is true ($p = 0$), if Γ is continuous at $\Gamma^{-1}(1 - \alpha)$:*

$$c_{n,b}(1 - \alpha) \xrightarrow{p_n} \Gamma^{-1}(1 - \alpha), \quad P_n(S_n > c_{n,b}(1 - \alpha)) \rightarrow \alpha.$$

2. *Under local alternative A2a ($p \neq 0$), if Γ is continuous at $\Gamma^{-1}(1 - \alpha)$:*

$$c_{n,b}(1 - \alpha) \xrightarrow{p_n} \Gamma^{-1}(1 - \alpha), \quad P_n(S_n > c_{n,b}(1 - \alpha)) \rightarrow \beta,$$

where $\beta = \Pr(f(v_0(\cdot) + p(\cdot)) > \Gamma^{-1}(1 - \alpha))$.

3. *Under global alternative A2b, $S_n \xrightarrow{p_n} \infty$:*

$$c_{n,b}(1 - \alpha) = O_p(1), \quad P_n(S_n > c_{n,b}(1 - \alpha)) \rightarrow 1.$$

4. $\Gamma(x)$ is absolutely continuous at $x > 0$ when the covariance function of v is nondegenerate a.e. in τ .

Thus the resampling mechanism consistently estimates the critical values, and the resampling tests are asymptotically unbiased and have the *same power* as the corresponding test in Theorem 2 that uses a known critical value. Thus, the test are consistent and have non-trivial power against the $1/\sqrt{n}$ -local alternatives, as discussed after Theorem 2.

4.5. Practical Implementation. Here we discuss practical implementation of the resampling inference.

4.5.1. Discretization. It is more practical to use a grid \mathcal{T}_n in place of \mathcal{T} with the largest cell size $\delta_n \rightarrow 0$ as $n \rightarrow \infty$.

Corollary 5. *Theorems 1-3 are valid for piece-wise constant approximations of the finite-sample processes using \mathcal{T}_n , given that $\delta_n \rightarrow 0$ as $n \rightarrow \infty$.*

4.5.2. *Choice and Estimation of $\Lambda(\tau)$, $J(\tau)$ $H(\tau)$.* In order to increase the test's power we could set

$$\Lambda^*(\tau) = [\Omega^*(\tau)]^{-1} = \text{Var}[z_i(\tau)]^{-1},$$

which is a (generalized) Anderson-Darling weight.⁹ In iid samples, there are many methods for estimating $\Lambda^*(\tau)$, *uniformly consistently in τ* .

An obvious and uniformly consistent by I3-4 estimate of $\Omega^*(\tau)$ is given by:

$$\begin{aligned}\hat{\Omega}^*(\tau) &= \frac{1}{n} \sum_{i=1}^n \hat{z}_i(\tau) \hat{z}_i(\tau)' \\ \hat{z}_i(\tau) &= R(\tau) \left[\hat{J}(\tau)^{-1} l_i(\tau, \hat{\theta}(\tau)) \hat{\Psi}_i(\tau) - \hat{H}(\tau)^{-1} d_i(\tau, \hat{\tau}(\tau)) \hat{\Upsilon}_i(\tau) \right].\end{aligned}$$

A uniformly consistent estimate of $J(\tau)$ is given by Powell's (1986) estimator,

$$\hat{J}(\tau) = \frac{1}{n} \sum_{i=1}^n K_h(Y_i - D_i' \hat{\alpha}(\tau) - X_i' \hat{\beta}(\tau)) \hat{\Psi}_i(\tau) [D_i, X_i],$$

where $K_h(x) = h^{-1}K(x/h)$ and $K(\cdot)$ is a compactly supported symmetric kernel with two uniformly bounded derivatives, and $h \sim Cn^{-1/5}$. Estimates of H are only needed in Examples 2 and 4. In Example 2,

$$\hat{H}(\tau) = \hat{J}(\tfrac{1}{2}),$$

and in Example 4, a uniformly consistent estimate of $H(\tau)$ is given by Powell's estimator:

$$\hat{H}(\tau) = \frac{1}{n} \sum_{i=1}^n K_h(Y_i - \tilde{X}_i' \hat{\vartheta}(\tau)) \tilde{X}_i \tilde{X}_i', \quad \tilde{X}_i = (D_i', X_i').$$

4.5.3. *Choice of the Block Size.* In Politis, Romano, and Wolf (1999) various rules are suggested for choosing an appropriate subsample size, including the calibration and minimum volatility methods. The calibration method involves picking the optimal block size and appropriate critical values on the basis of simulation experiments conducted with a model that approximates the situation at hand. The minimum volatility method involves picking (or combining) among the block sizes that yield the most stable critical values. More detailed suggestions emerge from Sakov and Bickel (1999) who suggest that choosing $b = kn^{2/5}$ yields optimal performance for a related subsampling method that involves recomputing the estimates. In our setting, this choice also appears reasonable. Our experiments and those in Chernozhukov (2002) indicate that selecting k between 3 and 10 are attractive both computationally and qualitatively.

⁹This choice is not readily suited to Example 2, since $\text{Var } z_i(\tfrac{1}{2}) = 0$. However, we can cut out $[\tfrac{1}{2} - \epsilon, \tfrac{1}{2} + \epsilon]$ from the interval \mathcal{T} . Alternatively, one may always simply use $\Lambda(\tau) = I$.

5. EMPIRICAL ANALYSIS

This section contains results from two empirical examples that illustrate the use of the inference procedures presented in this paper. The examples correspond to the models outlined in Sections 2.2 and 2.3. In the first example, we estimate the market demand for fish, and in the second, we consider the returns to schooling.

5.1. The Structure of Demand for Fish. In this section, we present estimates of demand elasticities which may potentially vary with the level of demand, τ , and assess the impact of price on the quantity distribution. The data contain observations on price and quantity of fresh whiting sold in the Fulton fish market in New York over the five month period from December 2, 1991 to May 8, 1992. These data were used previously in Graddy (1995) to test for imperfect competition in the market and later in Angrist and Imbens (2000) to illustrate the use of the conventional IV estimator as a weighted average of heterogeneous demands. The price and quantity data are aggregated by day, with the price measured as the average daily price for the dealer and the quantity as the total amount of fish sold that day. The data also contain information on the day of the week of each observation and variables indicating weather conditions at sea, which are used as instruments to identify the demand equation. The total sample consists of 111 observations for the days in which the market was open during the sample period.

The demand function we estimate takes a standard Cobb-Douglas form:

$$Q_{\ln(Y_p)|X}(\tau) = \lambda(\tau) + \alpha(\tau)\ln(p),$$

where Y_p is the demand when price is set at p . The elasticity $\alpha(\tau)$ is allowed to vary across the quantiles, τ , of the demand level. Note that this is a Cobb-Douglas version of the demand model with random elasticity discussed in Section 2.2.

The left panel of Figure 1 provides the estimates of elasticities obtained by IV-QR of $\ln(Y)$ on $\ln(P)$ using wind speed as the instrumental variable, while the right panel depicts standard quantile regression (QR) estimates of the effect of $\ln(P)$ on $\ln(Y)$. The shaded region in each figure represents the 90% confidence interval. The estimated model does not include covariates, but the estimated elasticities are not sensitive to the inclusion of controls for days of the week or other covariates.

In general, the point estimates obtained through IV-QR differ substantially from the corresponding QR estimates. The QR estimates appear to be approximately constant across the entire range of the quantity distribution and are uniformly small in magnitude. The IV-QR estimates, on the other hand, demonstrate a great deal of variability, ranging from near -2 at low quantiles to -0.5 in the upper end of the distribution. Except at high quantiles, the IV-QR estimates of the elasticities are uniformly greater in magnitude than the price effects predicted by QR, demonstrating an upward bias induced by the joint determination of price and quantity in the market. Also note that, like 2SLS and OLS, the interpretation

TABLE 1. The test results for Demand Equation, using $b = 50$ (subsampling with replacement)

Null Hypothesis	Alternative Hypothesis	P-value for CM Test Statistic
No Effect. $\alpha(\cdot) = 0$	Non-zero Effect	$< .01$
Fixed Elast. $\alpha(\cdot) = \alpha$	Random Elasticity	.28
Dominance $\alpha(\cdot) \leq 0$	Non-Dominance	.50
Exogeneity $\alpha(\cdot) = \alpha_{QR}(\cdot)$	Endogeneity	.18

of the IV-QR and QR estimates are very different. IV-QR estimates a demand model, while QR estimates the conditional quantiles of the equilibrium quantity as a function of the equilibrium price.

Results from formal tests of the location-shift hypothesis and the endogeneity hypotheses are contained in Table 1. We fail to reject the dominance hypothesis that $\alpha(\cdot) \leq 0$, indicating downward sloping demand at all quantiles. The rest of the results require careful discussion in view of the small sample $n = 111$. Due to the simultaneous determination of price and quantity, as expected there is clear evidence against the null of exogeneity. In particular, we reject the null of no-endogeneity at the 10% level (recall that variances cancel each other for Hausman tests) for the lower quartile ($\tau = .25$) of demand. However, the overall results are weaker, giving p-values of only 18%. The hypothesis of constant elasticity is also rejected at the 10% level when we compare the lower quartile elasticity with the median elasticity. The overall results are weaker yielding only a 28% p-value. These results indicate that the elasticity of demand is likely heterogeneous.

5.2. The Structure of the Returns to Schooling. As a further illustration of the use of the estimation and inference methods presented in this paper, we use the data and methodology employed in Angrist and Krueger (1991) to estimate the effects of schooling on earnings. In particular, we estimate linear conditional quantile models of the form

$$Q_{\ln(Y_s)|X}(\tau) = \alpha(\tau)S + X\beta(\tau),$$

where S is reported years of schooling and X is a vector of covariates, using quarter of birth as an instrument for education.¹⁰

We focus on the specification used in Angrist and Krueger (1991) which includes state of birth effects, year of birth effects, and a constant in the covariate vector.¹¹ The sample we consider consists of 329,509 males from the 1980 U.S. Census who were born between 1930 and 1939 and have data on weekly wages, years of completed education, state of birth,

¹⁰Specifically, we use the linear projection of S onto the covariates X and three dummies for first through third quarter of birth, with fourth quarter as the excluded category, as the instrumental variable.

¹¹Note that the estimates of the schooling coefficient are not sensitive to the specification of the X vector.

year of birth, and quarter of birth. The sample was selected based on the criteria found in Appendix 1 of Angrist and Krueger (1991).

IV-QR and QR estimates of the schooling coefficient are provided in Figure 2. The shaded region in each panel represents the 95% confidence interval. Both the quantile and inverse quantile regression estimates suggest that the “returns to schooling” vary over the earnings distribution. The first row in Table 2 reports the results from a test of the hypothesis of a constant treatment effect for the IV-QR estimates which is strongly rejected. The variability of the treatment effects is most apparent in the IV-QR estimates. While the QR estimates do vary statistically,¹² they are all closely clustered around the OLS estimate. The practical lack of variability in the QR estimates is clearly demonstrated in the first panel of Figure 2, which plots both the IV-QR (solid line) and QR (dashed line) estimates. Relative to the IV-QR estimates, the QR estimates appear to be approximately constant.

The shapes of the estimated treatment effects are also interesting. The QR estimates exhibit a distinct u-shape, implying higher returns to schooling to those in the tails of the distribution than to those in the middle. However, if schooling is endogenous to the earnings equation, these estimates do not consistently estimate the QTE and have no causal interpretation. IV-QR estimate, on the other hand, are consistent for the QTE under endogeneity and show quite different results than those obtained through standard QR. In particular, the IV-QR results show returns to schooling of approximately 20% per year of additional schooling at low quantiles in the earnings distribution which decrease as the quantile index increases toward the middle of the distribution and then remain approximately constant at levels near the QR and OLS estimates. This implies that the largest gains to additional years of schooling accrue to those at the low end of the earnings distribution. This observation is consistent with the notion that people with high unobserved “ability”, measured as the quantile index τ , will generate high earnings regardless of their education level, while those with lower “ability” gain more from the training provided by formal education.¹³

The third row of Table 2 reports the results from the test of stochastic dominance. As would be expected, it fails to reject the null hypothesis of stochastic dominance, confirming our intuition that schooling weakly increases earnings across the distribution.

In the final row of Table 2, we test the endogeneity hypothesis. The test strongly rejects the null hypothesis of no endogeneity, confirming the need to instrument for schooling in the earnings equation. Note that this result contrasts the conclusion which could be drawn from a Hausman test based on the 2SLS and OLS estimates, which fails to reject the null at the 5% level.¹⁴ Again, this confirms our intuition that endogeneity contaminates standard

¹²The test for the quantile regression estimates is not reported, but also strongly rejects the null hypothesis of a constant treatment effect.

¹³The term “ability” is used to characterize the unobserved component of earnings, which likely captures elements of ability and motivation as well as noise.

¹⁴The test statistic is 3.64 and is distributed as a χ^2_1 .

TABLE 2. The test results for Demand Equation, using $b = 1000$ (subsampling with replacement)

Null Hypothesis	Alternative Hypothesis	P-value for KS Test Statistic
No Effect. $\alpha(\cdot) = 0$	Non-zero Effect	$< .01$
Constant Effect. $\alpha(\cdot) = \alpha$	Varying Effect	.03
Dominance $\alpha(\cdot) \leq 0$	Non-Dominance	.50
Exogeneity $\alpha(\cdot) = \alpha_{QR}(\cdot)$	Endogeneity	.04

estimates of the returns to schooling and underscores the importance of accounting for this endogeneity in estimation.

Overall, the results indicate that the effect of schooling on earnings is quite heterogeneous, with the largest returns accruing to those who fall in the lower tail of the earnings distribution. The example also illustrates the variety of interesting distributional hypotheses that can be tested using the inference procedures presented in this paper. The results demonstrate that estimates of treatment effects which focus on a single feature of the outcome distribution may fail to capture the full impact of the treatment and that examining additional features may enhance our understanding of the economic relationships involved.

6. CONCLUSION

In this paper, we described how instrumental variable quantile regression can be used to evaluate the impact of treatment on the entire distribution of outcomes when the treatment is self-selected or selected in relation to potential outcomes. We introduced an instrumental variable quantile regression process and the set of inferences derived from it, focusing on tests of distributional equality, non-constant treatment effects, conditional dominance, and exogeneity. Inference, which is subject to the Durbin problem, was handled via a method of score resampling. In the paper, we demonstrated that the method is simple and computationally convenient and produces valid inference. The approach was illustrated through two examples: estimation of the demand curve in a supply-demand system and estimation of the returns to schooling. In both cases, the hypotheses of a constant treatment effect and exogeneity were rejected. The results suggest that estimates of treatment effects that focus on a single feature of the outcome distribution may fail to capture the full impact of the treatment and serve to illustrate the variety of distributional hypotheses that can be tested based on the quantile regression process.

APPENDIX A. PROOFS

We use the following empirical processes in the sequel, for $W \equiv (Y, D, X, Z)$

$$f \mapsto \mathbb{E}_n f(W) \equiv \frac{1}{n} \sum_{i=1}^n f(W_i), \quad f \mapsto \mathbb{G}_n f(W) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(W_i) - Ef(W_i)).$$

For example, if \hat{f} is estimated function, $\mathbb{G}_n f(W)$ means: $\frac{1}{\sqrt{n}} \sum_{i=1}^n (f(W_i) - Ef(W_i))_{f=\hat{f}}$. Outer and inner probabilities, P^* and P_* are defined as in van der Vaart (1998). In this paper \xrightarrow{P} means convergence in (outer) probability, and \xrightarrow{d} means convergence in distribution. $W_p \rightarrow 1$ means “with (inner) probability going to 1.” We will say that process $\{l \mapsto v_n(l), l \in \mathcal{L}\}$ is *stochastically equi-continuous* (s.e.) in $\ell^\infty(\mathcal{L})$ if for each $\epsilon > 0$ and $\eta > 0$, there is $\delta > 0$:

$$\limsup_{n \rightarrow \infty} P^* \left(\sup_{\rho(l, l') < \delta} |v_n(l) - v_n(l')| > \eta \right) < \epsilon$$

for some pseudo-metric ρ on \mathcal{L} , such that (\mathcal{L}, ρ) is totally bounded.

A.1. Proof of Proposition 1. ¹⁵ Conditioning on $X = x$ is suppressed. For P -a.e. value z of Z

$$\begin{aligned} & P[Y \leq q[D, \tau] | Z = z] \\ & \stackrel{(1)}{=} P[q[D, U_D] \leq q[D, \tau] | Z = z] \\ & \stackrel{(2)}{=} P[U_D \leq \tau | Z = z], \\ & \stackrel{(3)}{=} \int P[U_D \leq \tau | Z = z, V = v] dP[V = v | Z = z] \\ & \stackrel{(4)}{=} \int P[U_{\delta(z, v)} \leq \tau | Z = z, V = v] dP[V = v | Z = z] \\ & \stackrel{(5)}{=} \int P[U_o \leq \tau | Z = z, V = v] dP[V = v | Z = z] \\ & \stackrel{(6)}{=} P[U_o \leq \tau | Z = z] \\ & \stackrel{(7)}{=} \tau. \end{aligned} \tag{A.1}$$

Equality (1) is by A1 and A5. Equality (3) is by definition. Equality (4) is by A2. Equality (5) is by the similarity assumption A4: for each d , conditional on $(V = v, X = x, Z = z)$

$$U_{\delta(z, v)} \text{ equals in distribution to } U_o.$$

Equality (6) is by definition and equality (7) is by A3. Note that equality (2) is immediate when $\tau \mapsto q(d, \tau)$ is continuous, since we assumed that $\tau \mapsto q(d, \tau)$ is strictly increasing. To show (2) holds more generally, simply note that for $\tau \in (0, 1)$ the event $\{U_D \leq \tau\}$ implies the event $\{q[D, U_D] \leq q[D, \tau]\}$ by $\tau \mapsto q[d, \tau]$ non-decreasing on $(0, 1)$ for each d . On the other hand, the event $\{q[D, U_D] \leq q[D, \tau]\}$ implies the event $\{U_D \leq \tau\}$, since $\tau \mapsto q[d, \tau]$ is strictly-increasing and left-continuous¹⁶ in $(0, 1)$ for each d .

¹⁵The proof is that given in Chernozhukov and Hansen (2001) and is given here for completeness.

¹⁶ $\tau \mapsto q[d, \tau]$ is said to be left-continuous if $\lim_{\tau' \uparrow \tau} q[d, \tau'] = q[d, \tau]$.

Finally, since $\tau \mapsto q[d, \tau]$ is strictly increasing, left-continuous, we have

$$P[q[D, U_D] = q[D, \tau] | Z = z] = 0,$$

so that P -a.e.

$$P[Y \leq q[D, \tau] | Z] = P[Y < q[D, \tau] | Z]. \quad \blacksquare$$

A.2. Proof of Theorem 1. In the proof W denotes (Y, D, X, Z) . Define for $\vartheta \equiv (\beta, \gamma)$ and $\vartheta_0 \equiv (\beta(\tau), 0)$ and $\varphi_\tau(u) \equiv (1(u < 0) - \tau)$

$$\begin{aligned} \hat{f}(W, \alpha, \vartheta, \tau) &\equiv \varphi_\tau(Y - D'\alpha - X'\beta - \hat{\Phi}(\tau)'\gamma)\hat{\Psi}(\tau), \\ f(W, \alpha, \vartheta, \tau) &\equiv \varphi_\tau(Y - D'\alpha - X'\beta - \Phi(\tau)'\gamma)\Psi(\tau), \end{aligned}$$

where $\Psi(\tau) \equiv V(\tau) \cdot (X', \Phi(\tau))'$, $\Phi(\tau) \equiv \Phi(\tau, X, Z)$, $\hat{\Psi}(\tau) \equiv \hat{V}(\tau) \cdot (X', \hat{\Phi}(\tau))'$, $\hat{\Phi}(\tau) \equiv \hat{\Phi}(X, Z)$;

$$\begin{aligned} \hat{g}(W, \alpha, \vartheta, \tau) &\equiv \rho_\tau(Y - D'\alpha - X'\beta - \hat{\Phi}(\tau)'\gamma)\hat{V}(\tau), \\ g(W, \alpha, \vartheta, \tau) &\equiv \rho_\tau(Y - D'\alpha - X'\beta - \Phi(\tau)'\gamma)V(\tau), \end{aligned}$$

where $\rho_\tau(u) \equiv (\tau - 1(u < 0))u$. Let

$$Q_n(\alpha, \vartheta, \tau) \equiv \mathbb{E}_n \hat{g}(W, \alpha, \vartheta, \tau), \quad Q(\alpha, \vartheta, \tau) \equiv Eg(W, \alpha, \vartheta, \tau),$$

and

$$\begin{aligned} \hat{\vartheta}(\alpha, \tau) &\equiv (\hat{\beta}(\alpha, \tau), \hat{\gamma}(\alpha, \tau)) \equiv \arg \inf_{\vartheta \in \mathcal{B} \times \mathcal{G}} Q_n(\alpha, \vartheta), \\ \vartheta(\alpha, \tau) &\equiv (\beta(\alpha, \tau), \gamma(\alpha, \tau)) \equiv \arg \inf_{\vartheta \in \mathcal{B} \times \mathcal{G}} Q(\alpha, \vartheta, \tau), \\ \hat{\alpha}(\tau) &\equiv \arg \inf_{\alpha \in \mathcal{A}} \|\hat{\gamma}(\alpha, \tau)\|, \quad \alpha^* \equiv \arg \inf_{\alpha \in \mathcal{A}} \|\gamma(\alpha, \tau)\|. \end{aligned}$$

Step 1 (Identification) Here we show that $\theta(\tau) = (\alpha(\tau)', \beta(\tau)')$ uniquely solves the limit optimization problem. Define

$$\begin{aligned} \Pi(\theta, \tau) &\equiv E_P [\varphi_\tau(Y - D'\alpha - X'\beta)\Psi], \quad \Psi_i \equiv V_i \cdot [X_i' : \Phi_i']' \\ J(\theta, \tau) &= \frac{\partial}{\partial(\alpha', \beta')} E_P [\varphi_\tau(Y - D'\alpha - X'\beta)\Psi], \end{aligned}$$

We know by Proposition 1 that $\theta(\tau) = (\alpha(\tau)', \beta(\tau)')$ solves

$$\Pi(\theta(\tau), \tau) = 0.$$

Suppose there exists another θ in the parameter set that solves this equation

$$\Pi(\theta^*(\tau), \tau) = 0.$$

Then for any conformable non-zero vector λ Taylor expansion gives

$$\lambda' (\Pi(\theta^*(\tau), \tau) - \Pi(\theta, \tau)) = \lambda' J(\theta_\lambda(\tau), \tau) \lambda^* = 0$$

for $\theta_\lambda(\tau)$ is on the line segment that connects $\theta^*(\tau)$ and $\theta(\tau)$, and

$$\lambda^* = (\theta^*(\tau) - \theta(\tau))$$

which yields a contradiction by the full rank assumption after setting $\lambda = \lambda^*$.

So we have that, the true parameters $(\alpha(\tau), \beta(\tau))$ solve the equation

$$E\varphi_\tau(Y - D'\alpha(\tau) - X'\beta(\tau) - \Phi(\tau)'0)\Psi(\tau) = 0.$$

On the other hand, by R3 and by convexity in ϑ of the limit optimization problem for each τ and α , $\vartheta(\alpha, \tau)$ uniquely solves the equation:

$$E\varphi_\tau(Y - D'\alpha - X'\beta(\alpha, \tau) - \Phi(\tau)'\gamma(\alpha, \tau))\Psi(\tau) = 0.$$

We need to find $\alpha^*(\tau)$ such that this equation holds and the norm of $\gamma(\alpha, \tau)$ is as small as possible. $\alpha^* = \alpha(\tau)$ makes the norm of $\gamma(\alpha^*, \tau) = 0$ equal zero by Proposition 1. Thus $\alpha^*(\tau) = \alpha(\tau)$ is a solution; by the preceding argument it is unique and $\beta(\alpha^*(\tau), \tau) = \beta(\tau)$.

Step 2 (Consistency) By Lemma 2

$$\sup_{(\alpha, \vartheta, \tau) \in \mathcal{A} \times (\mathcal{B} \times \mathcal{G}) \times \mathcal{T}} \|Q_n(\alpha, \vartheta, \tau) - Q(\alpha, \vartheta, \tau)\| \xrightarrow{p} 0$$

This implies by Lemma 1 for extremum processes:

$$\sup_{(\alpha, \tau) \in \mathcal{A} \times \mathcal{T}} \|\hat{\vartheta}(\alpha, \tau) - \vartheta(\alpha, \tau)\| \xrightarrow{p} 0, \quad (\text{A.2})$$

which in turn implies

$$\sup_{(\alpha, \tau) \in \mathcal{A} \times \mathcal{T}} \|\|\hat{\gamma}(\alpha, \tau)\|_A - \|\gamma(\alpha, \tau)\|_A\| \xrightarrow{p} 0,$$

which by invoking Lemma 1 again implies

$$\sup_{\tau \in \mathcal{T}} \|\hat{\alpha}(\tau) - \alpha(\tau)\| \xrightarrow{p} 0,$$

which by equation (A.2) implies

$$\sup_{\tau \in \mathcal{T}} \|\hat{\beta}(\tau) - \beta(\tau)\| \xrightarrow{p} 0, \quad \sup_{\tau \in \mathcal{T}} \|\hat{\gamma}(\hat{\alpha}(\tau), \tau) - 0\| \xrightarrow{p} 0,$$

Step 3 (Asymptotics) By the computational properties of quantile regression estimator $\hat{\vartheta}(\alpha_n)$, for any $\alpha_n(\tau)$ in a small ball at $\alpha_\tau(\tau)$

$$O(K/\sqrt{n}) = \sqrt{n}\mathbb{E}_n \hat{f}(W, \alpha_n(\tau), \hat{\vartheta}(\alpha_n(\tau), \tau), \tau). \quad (\text{A.3})$$

By lemma 2, the following expansion of r.h.s. is valid for any $\alpha_n(\tau) - \alpha(\tau) \xrightarrow{p} 0$ uniformly in τ :¹⁷

$$\begin{aligned} \sqrt{n}\mathbb{E}_n \hat{f}(W, \alpha_n(\tau), \hat{\vartheta}(\alpha_n(\tau), \tau)) &\equiv \mathbb{G}_n \hat{f}(W, \alpha_n(\tau), \hat{\vartheta}(\alpha_n(\tau), \tau), \tau) \\ &\quad + \sqrt{n}E\hat{f}(W, \alpha_n(\tau), \hat{\vartheta}_n(\alpha_n(\tau), \tau), \tau) \\ &= \mathbb{G}_n f(W, \alpha(\tau), \vartheta(\alpha(\tau), \tau), \tau) + o_p(1) \\ &\quad + \sqrt{n}E\hat{f}(W, \alpha_n(\tau), \hat{\vartheta}(\alpha_n(\tau), \tau), \tau) \end{aligned} \quad (\text{A.4})$$

Expanding the last line further, uniformly in τ

$$\begin{aligned} O(K/\sqrt{n}) &= \mathbb{G}_n f(W, \alpha(\tau), \vartheta(\tau)) + o_p(1) \\ &\quad + (J_{\vartheta}(\tau) + o_p(1))\sqrt{n}(\hat{\vartheta}(\alpha_n(\tau), \tau) - \vartheta(\tau)) \\ &\quad + (J_{\alpha}(\tau) + o_p(1))\sqrt{n}(\alpha_n(\tau) - \alpha(\tau)), \end{aligned} \quad (\text{A.5})$$

¹⁷Note that by convention in empirical process theory $E\hat{f}(W)$ means $(Ef(W))_{f=\hat{f}}$.

where

$$J_{\vartheta}(\tau) = \frac{\partial}{\partial(\beta', \gamma')} E_P [\varphi_{\tau}(Y - D'\alpha(\tau) - X'\beta - \Phi(\tau)'\gamma)\Psi(\tau)]_{(\tau, \beta) = (0, \beta(\tau))},$$

$$J_{\alpha}(\tau) = \frac{\partial}{\partial(\alpha')} E_P [\varphi_{\tau}(Y - D'\alpha - X'\beta(\tau))\Psi(\tau)]_{\alpha = \alpha(\tau)},$$

In other words for any $\alpha_n(\tau) - \alpha(\tau) \xrightarrow{p} 0$ uniformly in τ

$$\begin{aligned} \sqrt{n}(\hat{\vartheta}(\alpha_n(\tau), \tau) - \vartheta(\tau)) &= -J_{\vartheta}^{-1}(\tau) \mathbb{G}_n f(W, \alpha(\tau), \vartheta(\tau)) \\ &\quad - J_{\vartheta}^{-1}(\tau) J_{\alpha}(\tau) [1 + o_p(1)] \sqrt{n}(\alpha_n(\tau) - \alpha(\tau)) + o_p(1), \text{ i.e} \\ \sqrt{n}(\hat{\gamma}(\alpha_n(\tau), \tau) - 0) &= -\bar{J}_{\gamma}(\tau) \mathbb{G}_n f(W, \alpha_{\tau}, \vartheta_{\tau}) \\ &\quad - \bar{J}_{\gamma}(\tau) J_{\alpha}(\tau) [1 + o_p(1)] \sqrt{n}(\alpha_n(\tau) - \alpha(\tau)) + o_p(1), \end{aligned}$$

where

$$[\bar{J}_{\beta}(\tau)' : \bar{J}_{\gamma}(\tau)']' \equiv J_{\vartheta}^{-1}(\tau).$$

Center a shrinking closed ball at $\alpha(\tau)$ for each τ and denote those balls $B_n(\alpha(\tau))$, wp $\rightarrow 1$,

$$\hat{\alpha}(\tau) = \arg \inf_{\alpha_n(\tau) \in B_n(\alpha(\tau))} \|\hat{\gamma}(\alpha_n(\tau), \tau)\|_A.$$

Observe that by Lemma 2

$$\sqrt{n}\|\hat{\gamma}(\alpha_n(\tau), \tau)\|_A = \|O_p(1) - \bar{J}_{\gamma}(\tau) J_{\alpha}(\tau) [1 + o_p(1)] \sqrt{n}(\alpha_n(\tau) - \alpha(\tau))\|_A,$$

Since $\bar{J}_{\gamma}(\tau) J_{\alpha}(\tau)$ and A have full rank, $\sqrt{n}(\hat{\alpha}(\tau) - \alpha(\tau)) = O_p(1)$. Hence by lemma 1

$$\sqrt{n}(\hat{\alpha}(\tau) - \alpha(\tau)) \stackrel{A}{=} \arg \inf_{\mu} \|\bar{J}_{\gamma}(\tau) \mathbb{G}_n f(W, \alpha(\tau), \vartheta(\tau)) - \bar{J}_{\gamma}(\tau) J_{\alpha}(\tau) \mu\|_A.$$

where $\stackrel{A}{=}$ means that the plims of the lhs and rhs agree in $\ell^{\infty}(\mathcal{T})$. Conclude that:

$$\begin{aligned} \sqrt{n}(\hat{\alpha}(\tau) - \alpha(\tau)) &\stackrel{A}{=} - \left(J_{\alpha}(\tau)' \bar{J}_{\gamma}(\tau)' A \bar{J}_{\gamma}(\tau) J_{\alpha}(\tau) \right)^{-1} \\ &\quad \times \left(J_{\alpha}(\tau)' \bar{J}_{\gamma}(\tau)' A \bar{J}_{\gamma}(\tau) \right) \mathbb{G}_n f(W, \alpha(\tau), \vartheta(\tau)) = O_p(1) \end{aligned}$$

and

$$\begin{aligned} \sqrt{n}(\hat{\vartheta}(\hat{\alpha}(\tau), \tau) - \vartheta(\alpha(\tau), \tau)) &\stackrel{A}{=} -J_{\vartheta}^{-1}(\tau) \left[I - J_{\alpha}(\tau) \left(J_{\alpha}(\tau)' \bar{J}_{\gamma}(\tau)' A \bar{J}_{\gamma}(\tau) J_{\alpha}(\tau) \right)^{-1} J_{\alpha}(\tau)' \bar{J}_{\gamma}(\tau)' A \bar{J}_{\gamma}(\tau) \right] \\ &\quad \times \mathbb{G}_n f(W, \alpha(\tau), \vartheta(\tau)) = O_p(1) \end{aligned}$$

Now note that due to simplifications because of invertibility of $J_{\alpha} \bar{J}_{\gamma}$ we have

$$\begin{aligned} \sqrt{n}(\hat{\gamma}(\hat{\alpha}(\tau), \tau) - \gamma(\alpha(\tau), \tau)) &\stackrel{A}{=} \left[-\bar{J}_{\gamma}(\tau) \left[I - J_{\alpha}(\tau) \left(J_{\alpha}(\tau)' \bar{J}_{\gamma}(\tau)' \right)^{-1} \bar{J}_{\gamma}(\tau) \right] \mathbb{G}_n f(W, \alpha(\tau), \vartheta(\tau)) \right. \\ &\quad \left. = 0 \times O_p(1) \right] \end{aligned}$$

Instead of working out the algebra to see a drastic simplification, using this fact and putting $(\alpha_n(\tau), \vartheta(\alpha_n(\tau), \tau)) = (\hat{\alpha}(\tau), \vartheta(\hat{\alpha}(\tau), \tau)) = \left(\hat{\alpha}(\tau), \hat{\beta}(\tau), 0 + o_p(1/\sqrt{n}) \right)$ back into the expansion (A.5) we have uniformly in τ

$$\mathbb{G}_n f(W, \alpha(\tau), \vartheta(\tau)) = J(\tau) \sqrt{n} \begin{pmatrix} \hat{\alpha}(\tau) - \alpha(\tau) \\ \hat{\beta}(\tau) - \beta(\tau) \end{pmatrix} + o_p(1)$$

Next by Lemma 2

$$\mathbb{G}_n f(W, \alpha(\tau), \vartheta(\tau)) \Rightarrow \mathbb{G}(\tau) \text{ in } \ell^\infty(\mathcal{T})$$

where $\mathbb{G}(\tau)$ is the Gaussian process with covariance function

$$S(\tau, \tau') = (\min(\tau, \tau') - \tau\tau')E\Psi(\tau)\Psi(\tau)'$$

which yields the desired conclusion

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}(\tau) - \alpha(\tau) \\ \hat{\beta}(\tau) - \beta(\tau) \end{pmatrix} \Rightarrow [J(\tau)]^{-1} \mathbb{G}(\tau). \quad \blacksquare$$

A.3. Proof of Theorem 2. Since f is either Kolmogorov-Smirnov or Cramer-von-Mises or one-sided version of these statistics, Part 1 follows by continuous mapping Theorem. Part 2 follows by observing that

$$f(\sqrt{n}g(\cdot)) \xrightarrow{p_n} \infty \implies f(\sqrt{n}g(\cdot) + G_n(\cdot)) \xrightarrow{p_n} \infty,$$

for any tight element $G_n(\cdot) = O_p(1)$ in $\ell^\infty(\mathcal{T})$ such f , once the null is violated (once the composite null is violated for one-sided tests). We also need that the distribution function of these limiting statistics is continuous. This follows by Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998): the distribution of functionals $f(v_0(\cdot) + p(\cdot))$ where f is of the specified sort, is absolutely continuous at $x > 0$ once $v_0(\cdot)$ has a nondegenerate covariance kernel.

That $\beta > \alpha$ follows from a generalized Andersen's Lemma for general Banach spaces, Lemma 3.11.4 in van der Vaart and Wellner (1996). \blacksquare

A.4. Proof of Proposition 2. Immediate from assumptions. \blacksquare

A.5. Proof of Theorem 3. To simplify the presentation, we assume that $\Lambda(\tau), J(\tau), H(\tau)$ are known. However, the case with the estimated matrices is straightforward by e.g. using the arguments in the proof Propostion 1 in Chernozhukov (2002) in part II of this proof.

We begin by showing Part 1 using the following steps.

Step I. By assumption realizations of function

$$\tau \mapsto \hat{z}(W, \tau)$$

belongs to a Donsker set of functions. We will denote this set as

$$\{\xi(W, \tau), \tau \in \mathcal{T}, \xi \in \Xi\}$$

Consider the empirical process

$$(\tau, \xi) \mapsto \mathbb{G}_n(\xi(\tau)),$$

which is Donsker by assumption with limit law denoted $J(P)$. Consider also its subsample realizations

$$(\tau, \xi) \mapsto \mathbb{G}_{j,b,n}(\xi(\tau)) \equiv \frac{1}{\sqrt{b}} \sum_{i \in J_i} (\xi(W_i, \tau) - E\xi(W_i, \tau)), \quad j = 1, \dots, B_n.$$

Let $J_n(P_n)$ denote the sampling distribution of $(\tau, \xi) \mapsto \mathbb{G}_n(\xi(\tau))$, and let $L_{b,n}$ denote the subsampling distribution of $(\tau, \xi) \mapsto \mathbb{G}_{j,b,n}(\xi(\tau))$. By Theorem 7.4.1 in Politis, Romano, and Wolf (1999)

$$\rho_L(J_n(P_n), L_{b,n}) \xrightarrow{P} 0 \text{ and } \rho_L(J(P), L_{b,n}) \xrightarrow{P} 0, \quad (\text{A.6})$$

where ρ_L denotes the Bounded-Lipschitz metric (Levy metric) that metrizes weak convergence.

Next let

- $J_n(P_n, \xi)$ denote the sampling (outer) law of $\tau \mapsto [\mathbb{G}_n(\xi(\tau))]$,
- $L_{b,n}(\xi)$ denote the subsampling (outer) law of $\tau \mapsto [\mathbb{G}_{j,b,n}(\xi(\tau))]$,
- $J(P, \xi)$ denote the limit law of $\tau \mapsto [\mathbb{G}_n(\xi(\tau))]$,

By (A.6) we have by definition of ρ_L

$$\sup_{\xi \in \Xi} [\rho_L(J_n(P_n, \xi), L_{b,n}(\xi))] \xrightarrow{P} 0 \text{ and } \sup_{\xi \in \Xi} [\rho_L(J(P, \xi), L_{b,n}(\xi))] \xrightarrow{P} 0,$$

since the projection maps $\tau \mapsto \mathbb{G}_n(\xi(\tau))$ are bounded uniform in ξ Lipschitz functionals of $(\xi, \tau) \mapsto \mathbb{G}_n(\xi(\tau))$. This means that

$$[\rho_L(J_n(P_n, \hat{z}), L_{b,n}(\hat{z}))] \xrightarrow{P} 0 \text{ and } [\rho_L(J(P, z), L_{b,n}(\hat{z}))] \xrightarrow{P} 0, \quad (\text{A.7})$$

provided

$$J_n(P_n, \hat{z}) \Rightarrow J(P, z).$$

The last observtaion follows from the assumed in I.4 (a) Donskerness, assumed orthogonality I.4 (b), and continuity with respect to the $L_2(P)$ semi-metric by I.4(a):

$$\sup_{\tau} \left\| E \left\| \mathbb{G}_n(\xi(\tau)) - \mathbb{G}_n(z(\tau)) \right\|^2 \right\|_{\xi=\hat{z}} \stackrel{A}{=} \sup_{\tau} \left| \text{Var} [\xi(W_i, \tau) - z(W_i, \tau)] \right|_{\xi=\hat{z}} \xrightarrow{P} 0,$$

For f denoting the two- and one-sided KS and CM functionals on the empirical processes, let

- H_n denote the (outer) distribution function of $f[\mathbb{G}_n(\hat{z}(\tau))]$,
- $H_{b,n}$ denote the subsampling distribution function of $f[\mathbb{G}_{j,b,n}(\hat{z}(\tau))]$,
- Γ denote the distribution function of $f[\mathbb{G}_{\infty}(z(\cdot))]$,

By (A.7) and definition of ρ_L

$$\rho_L(H_n, H_{b,n}) \xrightarrow{P} 0 \text{ and } \rho_L(H_n, H) \xrightarrow{P} 0,$$

since the functionals $f(\mathbb{G}_n(\xi(\cdot)))$ are bounded uniform Lipschitz functionals of $\tau \mapsto \mathbb{G}_n(\xi(\cdot))$.

Now we need to convert this result into convergence of distribution functions at continuity points. Γ is absolutely continuous (at $x > 0$ for one sided statistics) as shown in the Proof of Theorem 2. Since the statistics are real-valued, convergence with respect to Levy metric is equivalent to pointwise convergence at the continuity points x of $\Gamma(x)$

$$H_{b,n}(x) \xrightarrow{P} \Gamma(x).$$

Step II. Now note that we actually have the subsampling distribution $\Gamma_{b,n}$ of $f[v_{j,b,n}(\cdot)]$ and not $H_{b,n}$, but difference between $\Gamma_{b,n}$ and $H_{b,n}$ will be shown small. Indeed,

$$f[\mathbb{G}_{j,b,n}(\cdot)] - K_n \leq f[v_{j,b,n}(\cdot)] \leq f[\mathbb{G}_{j,b,n}(\cdot)] + K_n,$$

where e.g. when f is KS function:

$$\begin{aligned} K_n &= \sup_{\tau} \left\| \sqrt{b} \cdot [Ez(W, \tau)]_{z=\hat{z}} \right\|_{V(\tau)} \leq \sqrt{b} \cdot \sup_{\tau} \left\| C \|\hat{\vartheta}(\tau) - \vartheta(\tau)\| + C \|\hat{r}(\tau) - r(\tau)\| \right\| \\ &= O_p \left(\frac{\sqrt{b}}{\sqrt{n}} \right) = o_p(1) \end{aligned}$$

by invoking the condition I.4(b)-(c) and then I.3.

Thus $w_p \rightarrow 1$ $1(E_n) = 1$, where $E_n \equiv \{K_n \leq \delta\}$ for any $\delta > 0$.

Given the event E_n for a small $\epsilon > 0$ there is $\delta > 0$, $H_{b,n}(x - \epsilon)1(E_n) \leq \Gamma_{b,n}(x)1(E_n) \leq H_{b,n}(x + \epsilon)1(E_n)$ so that with probability tending to one:

$$H_{b,n}(x - \epsilon) \leq \Gamma_{b,n}(x) \leq H_{b,n}(x + \epsilon).$$

We have by Step I $H_{n,b}(x + c) \xrightarrow{p} \Gamma(x + c)$, for $c = \epsilon$ and $c = -\epsilon$, which implies

$$\Gamma(x - \epsilon) - \epsilon \leq \Gamma_{b,n}(x) \leq \Gamma(x + \epsilon) + \epsilon$$

w.p. $\rightarrow 1$. since ϵ can be set as small as we like and Γ is continuous at points of interest. This yields the conclusion $\Gamma_{b,n}(x) \xrightarrow{p} \Gamma(x)$.

Step III. Finally, convergence of quantiles is implied by the convergence of distribution functions at continuity points. E.g. Politis, Romano, and Wolf (1999).

Part 2 of Theorem 3 is immediate from Part 1 and contiguity.

Part 3 of Theorem 3 follows by steps that are identical to those in the proof of Part 1, except that we have convergence of subsampling distribution $\Gamma_{b,n}$ to some other distribution $\Gamma' \neq \Gamma$ at the continuity points. By tightness of Γ' , $c_n(1 - \alpha) = O_p(1)$ even if Γ' is not continuous at $\Gamma'^{-1}(1 - \alpha)$. Indeed, we have $c_n(1 - \alpha') \leq c_n(1 - \alpha) \leq c_n(1 - \alpha'')$, where α' and α'' are picked such that Γ' is continuous at some finite $\Gamma'^{-1}(1 - \alpha')$ and $\Gamma'^{-1}(1 - \alpha'')$ (possible by tightness). We then have by steps like in the proof of Part 1, $c_n(1 - \alpha') \xrightarrow{p} \Gamma'^{-1}(1 - \alpha')$ and $c_n(1 - \alpha'') \xrightarrow{p} \Gamma'^{-1}(1 - \alpha'')$.

Part 4 has already been proved in the proof of Theorem 2. ■

APPENDIX B. LEMMAS

Lemma 1 (Argmax Process). *Suppose that uniformly in π in a compact set Π and for a compact set K*

- i. $Z_n(\pi)$ is s.t. $Q_n(Z_n|\pi) \geq \sup_{z \in K} Q_n(z|\pi) - \epsilon_n$, $\epsilon_n \searrow 0$; $Z_n(\pi) = O_p(1)$ in $\ell^\infty(\Pi)$.
- ii. $Z_\infty(\pi) \equiv \operatorname{argsup}_{z \in K} Q_\infty(z|\pi)$ is uniquely defined continuous process.
- iii. $Q_n(\cdot|\cdot) \xrightarrow{p} Q_\infty(\cdot|\cdot)$ in $\ell^\infty(K \times \Pi)$, where $Q_\infty(\cdot|\cdot)$ is continuous. Then $Z_n(\cdot) \xrightarrow{p} Z_\infty(\cdot)$ in $\ell^\infty(\Pi)$.

Proof. The result is well known for the non-process case. The argument is just slightly more complicated than usual consistency arguments, cf. Amemiya (1985).

Suppose first that convergence in **iii.** holds uniformly in probability.

We have

$$Q_n(z|\pi) \xrightarrow{P} Q_\infty(z|\pi) \quad (\text{B.1})$$

where the convergence is uniform in (z, π) over compact sets. Uniformly in $\epsilon \in (c, c')$, $c' > c > 0$ wp $\rightarrow 1$, and uniformly in π : [i] $Q_n(Z_n(\pi)|\pi) \geq Q_n(Z_\infty(\pi)|\pi) - \epsilon/3$ by definition, [ii] $Q_\infty(Z_n(\pi)|\pi) > Q_n(Z_n(\pi)|\pi) - \epsilon/3$ by (B.1), [iii] $Q_n(Z_\infty(\pi)|\pi) > Q_\infty(Z_\infty(\pi)|\pi) - \epsilon/3$ by (B.1). Hence wp $\rightarrow 1$

$$Q_\infty(Z_n(\pi)|\pi) > Q_n(Z_n(\pi)|\pi) - \epsilon/3 \geq Q_n(Z_\infty(\pi)|\pi) - 2\epsilon/3 > Q_\infty(Z_\infty(\pi)|\pi) - \epsilon.$$

Pick any $\delta > 0$. Let $\{B(\pi), \pi \in \Pi\}$ be a collection of balls with diameter $\delta > 0$, each centered at $Z_\infty(\pi)$. Then $\epsilon \equiv \inf_{\pi \in \Pi} [Q_\infty(Z_\infty(\pi)|\pi) - \sup_{z \in K \setminus B(\pi)} Q_\infty(z)] > 0$ by assumption ii, and for any $\epsilon > 0$ we can pick c and c' so that for

$$P_*(\epsilon \in (c, c')) > 1 - \epsilon.$$

It now follows wp becoming greater than $1 - \epsilon$, uniformly in π

$$Q_\infty(Z_n(\pi)|\pi) > Q_\infty(Z_\infty(\pi)|\pi) - Q_\infty(Z_\infty(\pi)|\pi) + \sup_{z \in K \setminus B(\pi)} Q_\infty(z|\pi) = \sup_{z \in K \setminus B(\pi)} Q_\infty(z).$$

Thus wp becoming greater than $1 - \epsilon$,

$$\sup_{\pi \in \Pi} \|Z_n(\pi) - Z_\infty(\pi)\| \leq \delta.$$

But ϵ is arbitrary, so the preceding display occurs with probability converging to one. ■

Lemma 2 (Stochastic Expansions). *Under assumption R1-R4,*

I. *Uniformly in $(\alpha, \beta, \gamma, \tau)$ in $(\mathcal{A} \times \mathcal{B} \times \mathcal{G} \times \mathcal{T})$*

$$\mathbb{E}_n[\hat{g}(W, \alpha, \beta, \gamma, \tau)] \xrightarrow{P} E[g(W, \alpha, \beta, \gamma, \tau)].$$

II. *Uniformly in τ in \mathcal{T}*

$$\mathbb{G}_n \hat{f}(W, \hat{\alpha}(\tau), \hat{\beta}(\tau), \hat{\gamma}(\tau), \tau) = \mathbb{G}_n f(W, \alpha(\tau), \beta(\tau), 0, \tau) + o_p(1),$$

for any $(\hat{\alpha}(\tau), \hat{\beta}(\tau), \hat{\gamma}(\tau)) \xrightarrow{P} (\alpha(\tau), \beta(\tau), 0)$ uniformly in \mathcal{T} . Furthermore

$$\mathbb{G}_n f(W, \alpha(\tau), \beta(\tau), 0, \tau) \Rightarrow \mathbb{G}(\tau) \text{ in } \ell^\infty(\mathcal{T}),$$

where \mathbb{G} is a Gaussian process with covariance function $S(\tau, \tau')$ defined in Theorem 1.

Proof. We first show II. Denote $\pi = (\alpha, \beta, \gamma)$ and $\Pi = \mathcal{A} \times \mathcal{B} \times \mathcal{G}$ where \mathcal{G} is a closed ball at 0. We first show that the class of functions

$$\mathcal{H} \equiv \left\{ h = (\Phi, \Psi, \pi, \tau) \mapsto \varphi_\tau(Y - D'\alpha - X'\beta - \Phi(X, Z)'\gamma)\Psi(X, Z), \quad \pi \in \Pi, \Psi \in \mathcal{F}, \Phi \in \mathcal{F} \right\}$$

is Donsker, where \mathcal{F} is defined in R4.

The bracketing number of \mathcal{F} by Cor 2.7.4 in van der Vaart and Wellner (1996) satisfies

$$\log N_{[\cdot]}(\epsilon, \mathcal{F}, L_2(P)) = O\left(\frac{1}{\epsilon} \frac{\dim(z, \mathcal{Z})}{\eta}\right) = O\left(\frac{1}{\epsilon} 2^{2+\delta'}\right),$$

for some $\delta' < 0$. Thus \mathcal{F} is Donsker with a constant envelope. By Cor 2.7.4 in van der Vaart and Wellner (1996) the log of bracketing number of

$$\mathcal{X} \equiv \left\{ (\Phi, \pi) \mapsto (D'\alpha - X'\beta - \Phi(X, Z)'\gamma), \quad \pi \in \Pi, \Phi \in \mathcal{F} \right\}$$

satisfies

$$\log N_{[\cdot]}(\epsilon, \mathcal{X}, L_2(P)) = O\left(\frac{1}{\epsilon} \frac{\dim(z, d, \mathcal{Z})}{\eta}\right) = O\left(\frac{1}{\epsilon} 2^{2+\delta''}\right),$$

for some $\delta'' < 0$. Exploiting the monotonicity and boundedness of indicator function and assumption R3 the log of bracketing number of

$$\mathcal{V} \equiv \left\{ (\Phi, \pi) \mapsto 1(Y < D'\alpha + X'\beta + \Phi(X, Z)'\gamma), \quad \pi \in \Pi, \Phi \in \mathcal{F} \right\}$$

satisfies

$$\log N_{[\cdot]}(\epsilon, \mathcal{V}, L_2(P)) = O\left(\frac{1}{\epsilon} 2^{2+\delta''}\right),$$

as well. Therefore \mathcal{V} is Donsker since it has a constant envelope by R1 and R4.

Class \mathcal{H} is formed by taking products and sums of Donsker classes

$$\mathcal{F}, \mathcal{V}, \text{ and } \mathcal{T} = \{\tau \mapsto \tau\} :$$

$$\mathcal{H} = \mathcal{T} \cdot \mathcal{F} - \mathcal{V} \cdot \mathcal{F},$$

which is uniformly Lipschitz over $(\mathcal{T} \times \mathcal{F} \times \mathcal{V})$, and by Theorem 2.10.6 in van der Vaart and Wellner (1996) \mathcal{H} is Donsker.

Now we show (II) using the established Donskerness. Define the process

$$h = (\Phi, \Psi, \pi, \tau) \mapsto \mathbb{G}_n \varphi_\tau(Y - D'\alpha - X'\beta - \Phi(X, Z)'\gamma) \Psi(X, Z).$$

This process is Donsker (asymptotically Gaussian). Therefore the process

$$\tau \mapsto \mathbb{G}_n \varphi_\tau(Y - D'\alpha(\tau) - X'\beta(\tau)) \Psi(\tau, X, Z).$$

is also Donsker by linearity in τ and by the uniform Holder property of

$$\tau \mapsto (\alpha(\tau)', \beta(\tau)', \Phi(\tau, X, Z), \Psi(\tau, X, Z))$$

in τ with respect to the supremum norm, by R3 and R4. (To check the Donskerness, it is easy to verify (i) the definition of stochastic equicontinuity in τ with respect to the $L_2(P)$ semi-metric and (ii) finite-dimensional asymptotic normality by Linderberg-Levy CLT.) Thus we have

$$\mathbb{G}_n \varphi_\tau(Y - D'\alpha(\tau) - X'\beta(\tau)) \Psi(\tau, X, Z) \Rightarrow \mathbb{G}(\tau),$$

where $\mathbb{G}(\tau)$ has covariance function $S(\tau, \tau')$.

Since $\hat{\Psi}(\cdot) \xrightarrow{P} \Psi(\cdot)$ and $\hat{\Phi}(\cdot) \xrightarrow{P} \Phi(\cdot)$ uniformly over compacts and $\hat{\pi}(\tau) - \pi(\tau) \xrightarrow{P} 0$ uniformly in τ , we have by R3 and R4:

$$\delta_n \equiv \sup_{\tau \in \mathcal{T}} \rho(\hat{h}(\tau), h(\tau)) \xrightarrow{P} 0,$$

where ρ is the $L_2(P)$ semimetric on \mathcal{H} :

$$\rho(h, \tilde{h}) \equiv \text{Var} \left[\varphi_\tau(Y - D'\alpha - X'\beta - \Phi(X, Z)'\gamma) \Psi(X, Z) - \varphi_{\tilde{\tau}}(Y - D'\tilde{\alpha} - X'\tilde{\beta} - \tilde{\Phi}(X, Z)'\tilde{\gamma}) \tilde{\Psi}(X, Z) \right],$$

so that

$$\begin{aligned} & \sup_{\tau \in \mathcal{T}} \left\| \mathbb{G}_n \varphi_\tau(Y - D'\hat{\alpha}(\tau) - X'\hat{\beta}(\tau) - \hat{\Phi}(\tau, X, Z)'\hat{\gamma}(\tau)) \hat{\Psi}(\tau, X, Z) \right. \\ & \quad \left. - \mathbb{G}_n \varphi_\tau(Y - D'\alpha(\tau) - X'\beta(\tau)) \Psi(\tau, X, Z) \right\| \\ & \leq \sup_{\substack{\rho(\tilde{h}, h) \leq \delta_n \\ \tilde{h}, h \in \mathcal{H}}} \left\| \mathbb{G}_n \varphi_{\tilde{\tau}}(Y - D'\tilde{\alpha} - X'\tilde{\beta} - \tilde{\Phi}(X, Z)'\tilde{\gamma}) \tilde{\Psi}(X, Z) \right. \\ & \quad \left. - \mathbb{G}_n \varphi_\tau(Y - D'\alpha - X'\beta - \Phi(X, Z)'\gamma) \Psi(X, Z) \right\| \xrightarrow{P} 0 \end{aligned}$$

as $\delta_n \rightarrow 0$ by stochastic equicontinuity of $h \mapsto \mathbb{G}_n \varphi_\tau(Y - D'\alpha - X'\beta - \Phi(X, Z)'\gamma) \Psi(X, Z)$ (which is a part of being Donsker).

Having shown II, a simple way to show I. is to note that functions

$$\mathcal{P} = \{(\Phi, V, \alpha, \beta, \gamma, \tau) \mapsto \rho_\tau(Y - D'\alpha - X'\beta - \Phi(X, Z)'\gamma) V(X, Z)\}$$

are uniformly Lipschitz over

$$(\mathcal{F} \times \mathcal{F} \times \mathcal{A} \times \mathcal{B} \times \mathcal{G} \times \mathcal{T})$$

which by Theorem 2.10.6 in van der Vaart and Wellner (1996) and assumption R1 means that \mathcal{P} is Donsker. From this we have a uniform LLN

$$\begin{aligned} & \sup_{h \in \mathcal{H}} \left\| \mathbb{E}_n \rho_\tau(Y - D'\alpha - X'\beta - \Phi(X, Z)'\gamma) V(X, Z) \right. \\ & \quad \left. - E \rho_\tau(Y - D'\alpha - X'\beta - \Phi(X, Z)'\gamma) V(X, Z) \right\| \xrightarrow{P} 0 \end{aligned}$$

and by uniform consistency of $\hat{\Phi}$ and \hat{V} and R4 we have

$$\begin{aligned} & \sup_{(\alpha, \beta, \gamma, \tau) \in (\mathcal{A} \times \mathcal{B} \times \mathcal{G} \times \mathcal{T})} \left\| E \rho_\tau(Y - D'\alpha - X'\beta - \tilde{\Phi}(\tau, X, Z)'\tilde{\gamma}) \tilde{V}(\tau, X, Z) \right. \\ & \quad \left. - E \rho_\tau(Y - D'\alpha - X'\beta - \Phi(\tau, X, Z)'\gamma) V(\tau, X, Z) \right\| \xrightarrow{P} 0. \quad \blacksquare. \end{aligned}$$

APPENDIX C. VERIFICATION OF LINEAR REPRESENTATIONS

Proposition 3. *The conditions I.3 and I.4 are verified for the proposed implementation in Examples 1-4 under conditions R.1-R4.*

Proof. In Example 1, in the test of equality of distributions, I.3 is satisfied for $\theta(\cdot)$ by Theorem 1 by contiguity of P_n relative to P under which Theorem 1 was proven. Since $r = 0$,

$$z_i(\tau) = R(\tau) [J(\tau)^{-1} l_i(\tau, \theta(\tau)) \Psi_i(\tau)],$$

where

$$l_i(\tau, \theta(\tau)) = (\tau - 1(Y_i < D_i \alpha(\tau) + X_i' \beta(\tau))), \Psi_i(\tau) = V_i(\tau) [\Phi_i(\tau)', X_i']' \quad (\text{C.1})$$

Condition I.4(a) is checked in the proof of Lemma 2 in Appendix B, cf. the class of functions \mathcal{H} . Condition I.4.(b) holds by Proposition 1 and since Ψ_i is a function of (X_i, Z_i) only. Condition I.4(c) holds by the bounded density condition R.3.

In Example 2, in the test of constant effect, $\hat{\tau}(\cdot) = \hat{\theta}(\frac{1}{2})$ is an IV-QR estimate. Thus for $l_i(\cdot)$ defined in (C.1)

$$z_i(\tau) = R(\tau) [J(\tau)^{-1} l_i(\tau, \theta(\tau)) \Psi_i(\tau) - J(\frac{1}{2})^{-1} l_i(\frac{1}{2}, \theta(\frac{1}{2})) \Psi_i(\frac{1}{2})],$$

i.e. $d_i(\tau, r(\tau)) \Upsilon_i(\tau) = l_i(\frac{1}{2}, \theta(\frac{1}{2})) \Psi_i(\frac{1}{2})$. Thus I.3-I.4 hold by the preceding argument.

In Example 3, the test of Stochastic Dominance, r is also known, so the situation is like that in Example 1, for l_i defined in (C.1)

$$z_i(\tau) = R(\tau) [J(\tau)^{-1} l_i(\tau, \theta(\tau)) \Psi_i(\tau)],$$

so I.3 and I.4 are verified.

In Example 4, in the test of no-endogeneity, the estimate of $\hat{\tau}(\tau)$ is given by the ordinary QR of Y on D, X , denoted as $\hat{\vartheta}(\tau)$. In this case under conditions R.1-R.4, the estimator $\hat{\vartheta}(\tau)$ satisfies I3, e.g. Portnoy (1991):

$$\begin{aligned} \sqrt{n} \left(\hat{\vartheta}(\cdot) - \vartheta_n(\cdot) \right) &= H(\cdot)^{-1} n^{-1/2} \sum_{i=1}^n d_i(\cdot, \vartheta_n(\cdot)) + o_{p_n}(1), \\ d_i(\tau, \vartheta(\tau)) &= (\tau - 1(Y_i < \tilde{X}_i' \vartheta(\tau))) \tilde{X}_i, \quad \tilde{X}_i = (D_i', X_i')' \\ H(\tau) &= E f_{Y|\tilde{X}}(\vartheta(\tau)' \tilde{X}) \tilde{X} \tilde{X}'. \end{aligned}$$

Thus the score is given by

$$z_i(\tau) = R(\tau) [J(\tau)^{-1} l_i(\tau, \theta(\tau)) \Psi_i(\tau) - H(\tau)^{-1} d_i(\tau, \vartheta(\tau))].$$

The conditions I.3 and I.4 for $l_i(\tau, \theta(\tau)) \Psi_i(\tau)$ are checked above. As for $d_i(\tau, \vartheta)$, the proof of Lemma 2 checks I.4.(a) (put \tilde{X}_i in place of Ψ_i and $\gamma = 0$). Note that $E d_i(\tau, \vartheta(\tau)) = 0$ from the definition of the quantile regression coefficient $\vartheta(\tau)$, so I.4.(b) is satisfied, and I.4.(c) holds by the bounded density condition R.3. ■

REFERENCES

- ABADIE, A. (1995): "Changes in Spanish Labor Income Structure During the 1980s: A quantile Regression Approach," CEMPRF Working Paper No. 9521.
- (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," JASA, forthcoming.
- ABADIE, A., J. ANGRIST, AND G. IMBENS (2001): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, p. forthcoming.
- AMEMIYA, T. (1977): "The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model," *Econometrica*, 45(4), 955–968.
- (1985): *Advanced Econometrics*. Harvard University Press.
- ANDREWS, D. (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle, and D. McFadden. North Holland.
- ANDREWS, D. W. K. (1995): "Nonparametric kernel estimation for semiparametric models," *Econometric Theory*, 11(3), 560–596.
- ANGRIST, J. D., G. K., AND G. W. IMBENS (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67(3), 499–527.

- ANGRIST, J. D., AND A. B. KRUEGER (1991): "Does compulsory school attendance affect schooling and earnings," *Quarterly Journal of Economics*, 106, 979–1014.
- ATKINSON, A. B. (1970): "On the Measurement of Inequality," *Journal of Economic Theory*, 2, 244–263.
- CHERNOZHUKOV, V. (2002): "A Resampling Approach to Inference on Quantile Regression Process," January, Working Paper of MIT Department of Economics (www.SSRN.COM).
- CHERNOZHUKOV, V., AND C. HANSEN (2001): "An IV Model of Quantile Treatment Effects," September, Working Paper of MIT Department of Economics (www.SSRN.COM).
- DAVYDOV, Y. A., M. A. LIFSHITS, AND N. V. SMORODINA (1998): *Local properties of distributions of stochastic functionals*. American Mathematical Society, Providence, RI, Translated from the 1995 Russian original by V. E. Nazaikinskiĭ and M. A. Shishkova.
- DOKSUM, K. (1974): "Empirical probability plots and statistical inference for nonlinear models in the two-sample case," *Ann. Statist.*, 2, 267–277.
- FOSTER, J. E., AND A. F. SHORROCKS (1988): "Poverty orderings," *Econometrica*, 56(1), 173–177.
- GRADDY, K. (1995): "Testing for Imperfect Competition at the Fulton Fish Market," *Rand Journal of Economics*, 26(1), 75–92.
- HAUSMAN, J. A. (1978): "Specification tests in econometrics," *Econometrica*, 46(6), 1251–1271.
- HECKMAN, J. (1990): "Varieties of Selection Bias," *American Economic Review, Papers and Proceedings*, 80, 313–338.
- HECKMAN, J., AND R. ROBB (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in *Drawing Inference from Self-Selected Samples*, ed. by H. Wainer, pp. 63–107. Springer-Verlag, New York.
- HECKMAN, J. J., AND J. SMITH (1997): "Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts," *Rev. Econom. Stud.*, 64(4), 487–535. With the assistance of Nancy Clements, Evaluation of training and other social programmes (Madrid, 1993).
- HOGG, R. V. (1975): "Estimates of Percentile Regression Lines Using Salary Data," *Journal of the American Statistical Association*, 70(349), 56–59.
- HONG, H., AND E. TAMER (2001): "Estimation of Censored Regression with Endogeneity," Preprint.
- IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.
- KOENKER, R., AND G. S. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.
- KOENKER, R., AND Z. XIAO (2002): "Inference on the Quantile Regression Process," *Econometrica*, forthcoming.
- MACURDY, T., AND C. TIMMINS (1998): "Smoothed Quantile Estimation," Stanford, mimeo.
- MCFADDEN, D. (1989): "Testing for Stochastic Dominance," in *Studies in Economics of Uncertainty in Honor of Josef Hadar*, ed. by T. Fomny, and T. Seo.
- NEWKEY, W. K. (1990): "Efficient instrumental variables estimation of nonlinear models," *Econometrica*, 58(4), 809–837.
- (1997): "Convergence rates and asymptotic normality for series estimators," *J. Econometrics*, 79(1), 147–168.
- NEWKEY, W. K., AND J. L. POWELL (1990): "Efficient estimation of linear and type I censored regression models under conditional quantile restrictions," *Econometric Theory*, 6(3), 295–317.
- POLITIS, D. N., J. P. ROMANO, AND M. WOLF (1999): *Subsampling*. Springer-Verlag, New York.
- PORTNOY, S. (1991): "Asymptotic behavior of regression quantiles in nonstationary, dependent cases," *J. Multivariate Anal.*, 38(1), 100–113.
- POWELL, J. L. (1986): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.
- SAKOV, A., AND P. BICKEL (1999): "On the choice of m in the m out of n bootstrap in estimation problems," preprint, UC Berkeley.
- SHORACK, G. R., AND J. A. WELLNER (1986): *Empirical processes with applications to statistics*. John Wiley & Sons Inc., New York.
- VAN DER VAART, A. W. (1998): *Asymptotic statistics*. Cambridge University Press, Cambridge.
- VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak convergence and empirical processes*. Springer-Verlag, New York.

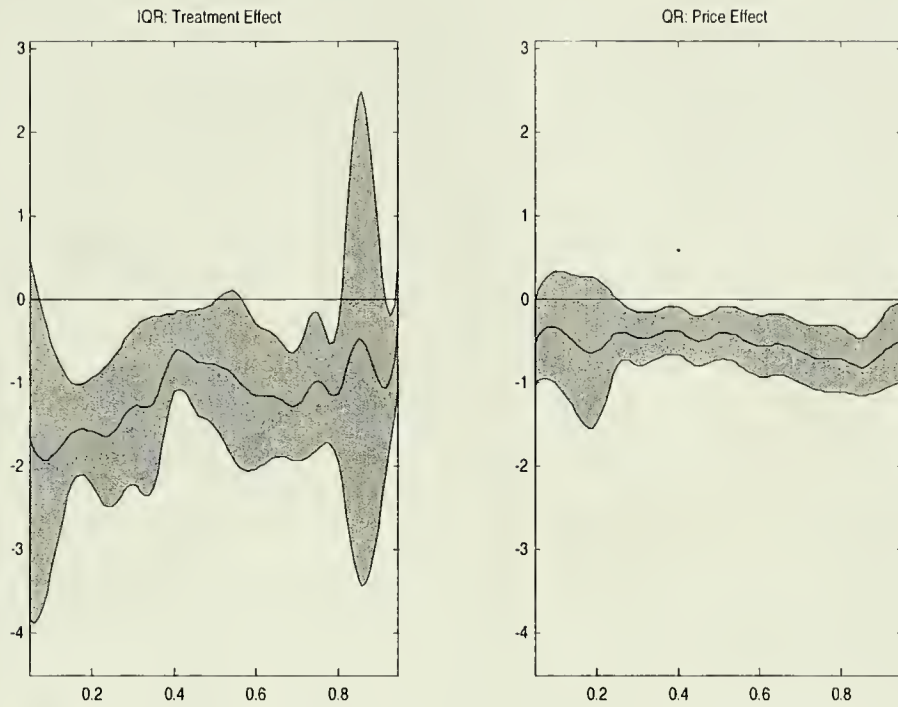


FIGURE 1. The sample size is 111. Coefficient estimates are on the vertical axis, while the quantile index is on the horizontal axis. The shaded region is the 90% confidence band estimated using robust standard errors. The first panel contains estimates of the price elasticity of demand obtained through instrumental variables quantile regression. The second panel presents estimates of the effect of $\ln(P)$ on $\ln(Q)$ obtained through standard quantile regression. Estimates were computed for $\tau \in [.05, .95]$.

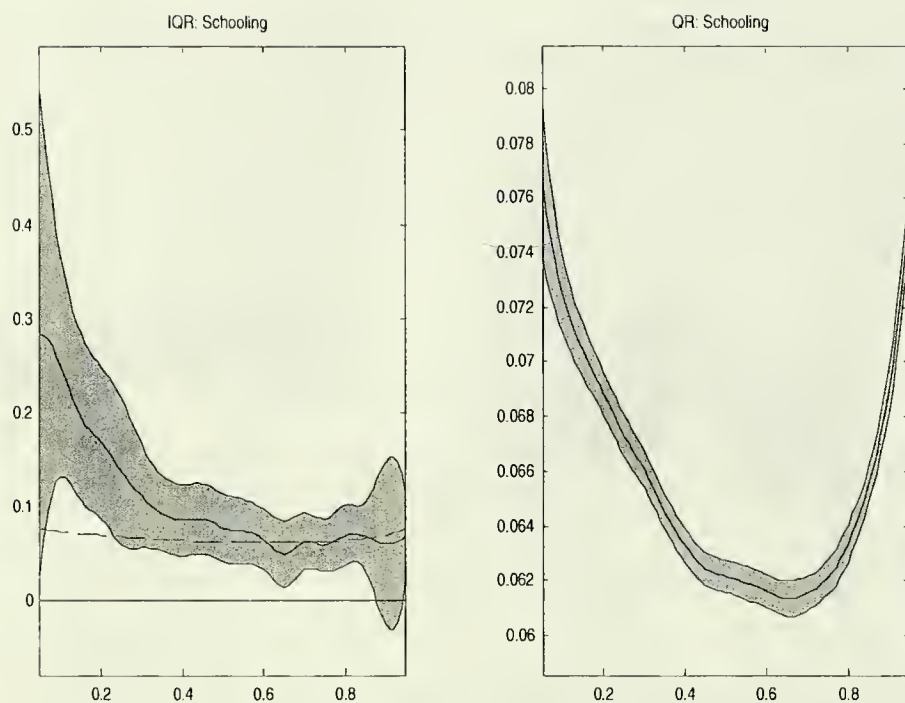


FIGURE 2. The sample size is 329,509. Coefficient estimates are on the vertical axis, while the quantile index is on the horizontal axis. The shaded region is the 95% confidence band estimated using robust standard errors. The first panel contains estimates of the returns to schooling obtained through instrumental variables quantile regression. The second panel presents estimates of the effect of years of schooling on earnings obtained through standard quantile regression. For comparison, the dashed line in the first panel plots the schooling coefficient estimated through standard quantile regression. All estimates were computed for $\tau \in [.05, .95]$.

MAR

2003

Date Due

Lib-26-67



3 9080 02246 2425

